

Run on: July 25, 2002, 05:58:15 ; Search time 2251.99 Seconds
(without alignments)

12849.734 million cell updates/sec

Title: US-09-821-839-1
Perfect score: 2144

Sequence: 1 actgcatcagccactctt.....aaaaaaaaaaaaaaaa 214

Scoring table:	IDENTITY_NUC	Canext 1 0
Canext 10 0		

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : RST: *

- 1: em_estha:*
- 2: em_esthm:*
- 3: em_estln:*
- 4: em_estum:*
- 5: em_estov:*
- 6: em_estol:*
- 7: em_estro:*
- 8: em_htc:*
- 9: gb_est1:*
- 10: gb_est2:*
- 11: gb_htc:*
- 12: gb_gss:*
- 13: gb_gss_hum:*
- 14: em_gss_inv:*
- 15: em_gss_pln:*
- 16: em_gss_vrt:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	609.8	28.4	1233	12	B10183	B10183 F19M12-Sp6
C 2	188.8	8.8	192	12	B97951	B97951 F2C23PHE I
C 3	188.8	8.8	818	12	BH530370	BH530370 BCGDDG70TFR
C 4	175	8.2	531	10	BF291837	BF291837 WMEZ304.F.E
C 5	172.6	8.1	166	12	B15531	B15531 F1C14-F7 IG
C 6	165.2	7.1	803	12	BH530380	BH530380 BCGDDG70TFR
C 7	152.4	7.1	465	12	BH462012	BH462012 BOGNK43TFR
C 8	147.4	6.9	727	12	BH487741	BH487741 BCGWBA48TFR
C 9	116.8	5.4	336	12	CMSO0DMB	AL080416 AbLabidops
C 10	96.4	4.5	498	10	BM270448	BM270448 sak13a03.
C 11	93.2	4.3	612	12	BH479564	BH479564 DOHKJ17TFR
C 12	80.8	3.8	605	12	BH462017	BH462017 BOHNK43TFR
C 13	70.6	3.3	791	12	AZ686821	AZ686821 ENTLLI74TFR
C 14	70.4	3.3	668	9	AL514901	AL514901 AE154901
C 15	68.8	3.2	1235	12	CNS0161D	AL106171 Drosophilid
C 16	68.4	3.2	1309	10	BE20736	BE20736 HMM002.AG
C 17	68	3.2	1101	12	CNS001EB	AL060732 Drosophila

C	18	67.8	3.2	330	9	AL513817	AL513817
C	19	67.8	3.2	634	9	AL514047	AL514047
C	20	67.6	3.2	410	9	AL513713	AL513713
C	21	67.6	3.2	590	9	AL515821	AL515821
C	22	67.4	3.1	423	9	AM119922	AM119922
C	23	67.2	3.1	613	12	A0922590	A0922590
C	24	67	3.1	661	12	CNS04Q0H	AL303074
C	25	66.2	3.1	987	12	CNS014PQ	AL104456
C	26	65.8	3.1	388	12	CNS039K1	AL234027
C	27	65.8	3.1	420	9	AL515191	AL515191
C	28	65.6	3.1	329	9	AL513719	AL513719
C	29	65.6	3.1	456	9	AL513813	AL513913
C	30	65.6	3.1	523	9	AL514015	AL514015
C	31	65.6	3.1	619	9	AL514935	AL514935
C	32	65.4	3.1	329	9	AL513719	AL513719
C	33	65.4	3.1	421	10	BM270739	BM270739
C	34	65.2	3.0	714	12	A0324694	AC324694
C	35	65.2	3.0	777	12	CNS025WB	AL182613
C	36	65.2	3.0	958	12	CNS0072R	AL066743
C	37	65	3.0	250	10	B174555	B174555
C	38	64.6	3.0	230	9	AL181812	AL181812
C	39	64.2	3.0	453	10	BF296095	BF296095
C	40	64.2	3.0	618	10	BE874847	BE874847
C	41	64.2	3.0	767	12	CNS00AQC	AL059524
C	42	64.2	3.0	951	12	A2672893	A2672893
C	43	64	3.0	954	12	BH162327	BH162327
C	44	63.8	3.0	279	9	AT540752	AT540752
C	45	63.8	3.0	633	9	AL513979	AL513979

ALIGNMENTS

RESULT 1

LOCUS	1233 bp	DNA	linear	GSS 14-MAY-1999
B10183				
DEFINITION	F19N12-Sp6 IGF Arabidopsis thaliana genomic clone F19N12, DNA			

VERSION	B10183.1	GI:2091302
---------	----------	------------

SOURCE	thale cress.
--------	--------------

Eukaryota; Viridipla

Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 1222)

AUTHORS Feng, J., Dewar, K., Buehler, E., Kim, C., Li, Y., Shinn, P., Sun, H. and Eckert, J.

TITLE BAC End Sequences at ATGC

COMMENT
other_GSSS: F19N12-1/
Contact: Ecker J.

Aradiopsis Indiana genome center
University of Pennsylvania

Dept: 01 Biology, University of Pennsylvania/ Philadelphia, PA 19104

Fax: 215-898-8780

Seq primer: Sp6

High quality sequence start: 87

FEATURES	Location/Qualifiers
1.000	

```
/db_xref="taxon:3702"
```

```
/clone_lib="IGF"
```

```
/note="Vector: BelobACII; Site_1: EcoRI; Site_2: EcoRI;
```


DEFINITION	BOGDQ70TF BOGD Brassica oleracea genomic clone BOGDQ70, DNA sequence.
ACCESSION	BH530370
VERSION	BH530370.1 GI:17746915
KEYWORDS	GSS.
SOURCE	Brassica oleracea.
ORGANISM	Brassica oleracea
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica. 1 (bases 1 to 818) Town,C.D., Van Aken,S., Uteirback,T. and Fraser,C.M. Whole genome shotgun sequencing of Brassica oleracea unpublished (2001)
AUTHORS	Other_GSSs: BOGDQ70TR
TITLE	Contact: Chris Town
JOURNAL	TIGR
COMMENT	9712 Medical Center Drive, Rockville, MD 20850, USA. Tel: 301-838-3523 Fax: 301-838-0208 Email: cdtown@tigr.org DNA is from a doubled haploid provided by Tom Osborn. Seq primer: TF Class: sheared ends. Location/Qualifiers 1..818 /organism="Brassica oleracea" /strain="T0100DHS" /db_xref="taxon:3712" /clone="BOGDQ70" /clone_1lb="BOGD" /note="Vector: pHOSt; Site:1; BstXI: 2-3 Kb sheared genomic DNA inserted into pHOSt using BstXI linkers"
BASE COUNT	264 a 186 c 168 g 200 t
ORIGIN	
Query Match	8.8%; Score 188.8; DB 12; Length 818;
Best Local Similarity	74.9%; Pred.No.2.3e-16;
Matches 352:	Conservative 0; Mismatches 87; Indels 31; Gaps 8;
OY	32 aaagaattccatcttcataatcgaattccaattcagttcaagtctcgtaac-- 89
DB	360 AACGAAGCTTCATTTCCGATTGGAAA--TCGGAATTCGAATCAACTAAACTTTCTTCGGAG 417
OY	90 -----gagaaanaatgaagagagatcgagatgaggaaatccaagcgactgagc 140
DB	418 AAGTTTCGTGTACACAAGATGAAGAATGCGTGCGAGATTTCMAAGCCCAAGCGCTGACC 477
OY	141 cgagcgccgttcgcgggaagaagctcgcgttcgacgcgatcagcgatcagcggaagaagacaga 200
DB	478 CGACGCCGTTCCCCGGGAAGAACCTTCGATCGATTTCATCTCCCGGAAGAGAGCTCAAA 537
OY	201 tcctccgcgttccttgatcatcacctctctggaacaanaaatcggagctctgctgctt 260
DB	538 TATCTCCAGATTCTATCCAGCACACCTCTCT-GCCAAGGAAGTAGAGATATCCGCTGCTT 596
OY	261 ctgtcgatctcgtctcggatgtgt-----agctgatatcaacgcttctctgtgtgtcga 314
DB	597 CCCTTATTTCTCCTCCCATTTTGCTCTTCGCACATCCAGCAGCAACGTTTTATCGGGTTCTA 656
OY	315 gagagatcgagaagagctgaaatccggaagaagactaaattgaagaggtagaagttctta 374
DB	657 GCAGAGTCCAGAAAGACT-----TGGGGAAGATATGAATTTGAAGAGGAAGAGATTCTTC 710
OY	375 aaactgttaaatagt---gaaggaagatgttgttgattcgaattcgaagattcga 431
DB	711 GACCTGGTAAGATGTAAAGACAGACACATCCTGATCCG-AAAGTTCCGAGGATCATCTA 769
OY	432 ggtcttaacttaagctacacaagaagagaagagatgatgatcgaaatga 481
DB	770 GATCATTAAGCTTAAGCTTAC-CAAAGGAAGATGTCAATGAATGGAGTGAAGCA 818

RESULT	4	531 bp	mRNA	linear	EST 17-NOV-2000
LOCUS	BF291837				
DEFINITION	BF291837	531 bp	mRNA	linear	EST 17-NOV-2000
ACCESSION	WHE2204.E10.J20S	Aegilops speltioides	another cDNA library	Aegilops	
VERSION	BF291837	speltioides cDNA clone WHE2204.E10.J20	mRNA sequence.		
KEYWORDS	BF291837.1	GI:11222901			
ORGANISM	EST.				
REFERENCE	Akhunov,E., Anderson,O.D., Chao,S., Chin,A., Choi,D.W., Close,T.J., J.				
AUTHORS	Fenton,R.D., Han,P.S., Hsia,C.C., Kang,T., Kianian,P., Lazlo,G.R.,				
	Miller,R., Otto,C., Rausch,C.J., Seaton,C.L., Simons,K., Tong,J.C.				
	and Zhang,D.				
TITLE	The structure and function of the expressed portion of the wheat				
JOURNAL	genomes - Another cDNA library from Aegilops speltioides				
COMMENT	Unpublished (2000)				
	Contact: Olin Anderson				
	US Department of Agriculture, Agriculture Research Service, Pacific				
	West Area, Western Regional Research Center				
	800 Buchanan Street, Albany, CA 94710, USA				
	Tel: 5105595773				
	Fax: 5105595818				
	Email: oanderson@pw.usda.gov				
	Sequence have been limited to remove vector sequence and low				
	quality sequence with phred score less than 20				
	Seq primer: Stratagene SK primer.				
FEATURES	Location/Qualifiers				
Source	1..531				
	/organism="Aegilops speltioides"				
	/cultivar="F2 from 2-12-4-8-1-1-1 (1) x PI36909-12-811-(1)"				
	/db_xref="taxon:4573"				
	/clone="WHE2204.E10.J20"				
	/clone_lib="Aegilops speltioides another cDNA library"				
	/tissue_type="Another"				
	/dev_stage="Premetotic anthers"				
	/lab_host="E. coli SOLR"				
	/note="Vector: Lambda Uni-ZAP XR, excised phagemid;				
	Site.1: EcoRI; Site.2: XhoI; plants were grown in a growth				
	chamber at the University of California, Davis (Akhunov)				
	Premetotic anthers were harvested, total RNA and poly(A)				
	RNA were prepared, from each tissue and then pooled, a				
	cDNA library was made, and the cDNA clones were in vivo				
	excised to give pluescript phagemids in the T7 Close lab				
	(Akhunov, Chin, Choi, Close, Fenton, Kianian, Otto, Simons,				
	Zhang) at the University of California, Riverside.				
	Plasmid DNA preparations and DNA sequencing were performed				
	in the OD Anderson lab (all other authors)."				
BASE COUNT	126 a 149 c 135 g 121 t				
ORIGIN					
Query Match	8.2%	Score 175;	DB 10;	Length 531;	
Best Local Similarity	61.5%	Pred. No. 1.8e-14;			
Matches 280; Conservative	0;	Mismatches 175;	Indels 0;	Gaps 0;	
1370	aaaagcgaagacacctaatctagctagtcggatgcgagcttactctgcgcacccaagaatt	1429			
4	AAAGGCTCTAGGAATGTGCAAGTTGCTGGGCATTTCCTGCATCACCCACCGGCATT	63			
1430	gaagaataatcaaccttacacagcaccgcggaaaaggaacttcaacattagaacctaga	1489			
64	GAAGAGAACCCACCATACATAGCATCAGCAAGAGTCTTCATGTGTAGGATCAACCTT	123			
1490	tattagccggcactgaagtgtgtgcacatggaggtgtgtgtttcaagaagtctcaactcaaa	1549			
124	TACAGCCGAGGAGGCTCGCTCCCATGAGAGTGCGCTGGTCCGAGGCTCTCGACTTCGAA	183			

QY 1081 ttcttgaaaccttcttcgcgaggaagaattcaactcctaagtgttaaggatgatgagga 1140
|| ||| | || | || | || ||||||||||||||||

QY 1581 -----ggttctaactaaagctgctgag 1604
|||||

Db 434 TGAAGATACATATACATCTTTTACTTCTCAGGTTTACTTAAAGCTGCTCAG 375

Qy 1605 ccaatccgaagttgaaggaagccaaatccttgctgctccatcctatccgaccaa 1664

Db 374 CTAATCGAAGAACTTGAAGGCTGCAATCCCTTGGCCGTCACGTCGATCACA 315

Qy 1665 ccaatcctgcttctggccctcaactgtagcagcagcagcctgctgctcgcctcagc 1724

Db 314 CACAACTCTGTTTGGCCCTCGAGTGTAGCAGCCGAGCTCTACTTCTGCTGCATCA 255

Qy 1725 aacacacacaaatcctcgtatccacaaagagtcataaagtcattgtagaaca 1780

Db 254 AACACAACAAATCTCAGCTTACCAAGAGTGTAAAGTACCACTCTTCAACA 199

RESULT 7

LOCUS BH462012 465 bp DNA linear GSS 13-DEC-2001

DEFINITION BOHNZ43TF BOHN Brassica oleracea genomic clone BOHNZ43, DNA sequence.

ACCESSION BH462012

VERSION BH462012.1 GI:17652941

KEYWORDS GSS.

SOURCE Brassica oleracea

ORGANISM Brassica oleracea

REFERENCE 1 (bases 1 to 465)

AUTHORS Town,C.D., Van Aken,S., Uterback,T. and Fraser,C.M.

TITLE Whole genome shotgun sequencing of Brassica oleracea

JOURNAL Unpublished (2001)

COMMENT Other_GSSs: BOHNZ43TR

CONTACT: Chris Town

TIGR 9712 Medical Center Drive, Rockville, MD 20850, USA.

Tel: 301-838-3523

Fax: 301-838-0208

Email: cdtown@tigr.org

DNA is from a doubled haploid provided by Tom Osborn.

Seq primer: TR

Class: sheared ends.

FEATURES

source 1..465

Location/Qualifiers

/organism="Brassica oleracea"

/strain="TO1000DH3"

/db_xref="taxon:3712"

/clone="BOHNZ43"

/note="Vector: PHOS1, Site 1: BstXI; 2-3 kb sheared genomic DNA inserted into PHOS1 using BstXI linkers"

BASE COUNT 126 a 70 c 116 g 153 t

ORIGIN

Query Match 7.1%; Score 152.4; DB 12; Length 465;

Best Local Similarity 88.7%; Pred. No. 2e-11;

Matches 105; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

Db 78 AGGTAC 73

RESULT 8

LOCUS BH487741 727 bp DNA linear GSS 13-DEC-2001

DEFINITION BOGWB48TR BOGW Brassica oleracea genomic clone BOGWB48, DNA sequence.

ACCESSION BH487741

VERSION BH487741.1 GI:17695845

KEYWORDS GSS.

SOURCE Brassica oleracea

ORGANISM Brassica oleracea

REFERENCE 1 (bases 1 to 727)

AUTHORS Town,C.D., Van Aken,S., Uterback,T. and Fraser,C.M.

TITLE Whole genome shotgun sequencing of Brassica oleracea

JOURNAL Unpublished (2001)

COMMENT Other_GSSs: BOGWB48TF

CONTACT: Chris Town

TIGR 9712 Medical Center Drive, Rockville, MD 20850, USA.

Tel: 301-838-3523

Fax: 301-838-0208

Email: cdtown@tigr.org

DNA is from a doubled haploid provided by Tom Osborn.

Seq primer: TR

Class: sheared ends.

FEATURES

source 1..727

Location/Qualifiers

/organism="Brassica oleracea"

/strain="TO1000DH3"

/db_xref="taxon:3712"

/clone="BOGWB48"

/note="Vector: PHOS1, Site 1: BstXI; 2-3 kb sheared genomic DNA inserted into PHOS1 using BstXI linkers"

BASE COUNT 108 a 168 c 196 g 175 t

ORIGIN

Query Match 6.9%; Score 147.4; DB 12; Length 727;

Best Local Similarity 67.9%; Pred. No. 8.1e-11;

Matches 324; Conservative 0; Mismatches 66; Indels 87; Gaps 5;

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 498)	Glycine. Shoemaker, R., Kelm, P., Vodka, L., Erpelting, J., Corvelli, V., Khanna, A., Bolla, B., Marra, M., Hillier, L., Kueba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Stepien, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Glibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.	Public Soybean EST Project Unpublished (1999)		
Public Soybean EST Project				
Contact: Shoemaker R/Public Soybean EST Project				
Public Soybean EST Project				
Washington University School of Medicine				
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA				
Tel: 314 286 1800				
Fax: 314 286 1810				
Email: est@watson.wustl.edu				
This clone is available through: Resgen, Invitrogen Corp. 2130				
South Memorial Parkway Huntville, AL 35801 For further information				
call: (800)-533-4363 or contact: cclures@resgen.com web site:				
www.resgen.com				
Seq primer: -40RP from Glbco				
High quality sequence stop: 422.				
Location/Qualifiers				
1. 498				
/organism="Glycine max"				
/db_xref="taxon:3847"				
/clone="SOYBEAN CLONE ID: Gm-c1075-217"				
/clone_lib="Gm-c1075"				
/tissue_type="differentiating somatic embryos cultured on				
MSM6AC"				
/lab_host="DH10B"				
/note="Vector: pBluescript II SK+ Site.1: EcoRI; Site.2:				
XhoI; The cDNA library was constructed from mRNA isolated				
from differentiating somatic embryos cultured on MSM6AC.				
The library was prepared using the Stratagene pBluescript				
II SK(+) library construction kit. Complementary DNA was				
synthesized from mRNA using a primer consisting of a				
poly(dT) sequence with an XhoI restriction site. EcoRI				
adaptors were ligated to the blunt-ended cDNA fragments				
followed by XhoI digestion. The cDNA fragments were				
directionally cloned into the EcoRI-XhoI restriction site				
of the pBluescript vector. The ligated cDNA fragments				
were transformed into E.coli Electromax DH10B host cells.				
Tissue culture and library construction were performed by				
Francoise Thibaud-Nissen and Anu Khanna (Lila Vodka lab,				
University of Illinois)."				
BASE COUNT	142 a	92 c	108 g	156 t
ORIGIN				
Query Match	4.5%	Score 96.4	DB 10	Length 498
Best Local Similarity	62.8%	Pred. No. 0.0061		
Matches 167	Conservative 0	Mismatches 96	Indels 3	Gaps 1
1572	actctgtgtgttctacttaaaagctgtctgagccaatccgaagttgtaaaggaagcca	1631		
64	aatgttttcaggtattacttaaaagcgttaatgctgacgcagaggttcgaagaagggcca	123		
1632	aatccttgctgttaccaccacatcgcagcaaatcgaactcgttcttggccccaatg	1691		
124	agtatctggcagagcgtgcacatgtagctgcaatgagcagctgtgcacggccttcaacag	183		
1692	tagcagctgcgaactcgtgtgtctcgcctcgcacgcagacaacaanaatctctgcataccaac	1751		
184	tttgctgcagcactgttaattctcgttcgttcttgaagtcataatcaaat--ttcattaccgcca	240		
1752	gagtcataaagaagtcacatctagaacaacagataagctgtccggaatgcgtttagagtc	1811		
241	aagtcattggagatcacttgcattgacataaagatgagaaattttgtacgaatgcattagagatc	300		
1812	tggaactggtctcgttggcgagtaagca	1837		

Db 301 TGGAGTGGGCTTACGATCTAGGA 326

RESULT 11

LOCUS BH477964 612 bp DNA linear GSS 13-DEC-2001

DEFINITION BOHKJ17R BOHK Brassica oleracea genomic clone BOHKJ17, DNA sequence.

ACCESSION BH477964

VERSION BH477964.1 GI:17686075

KEYWORDS GSS.

SOURCE Brassica oleracea.

ORGANISM Brassica oleracea

REFERENCE Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M. Whole genome shotgun sequencing of Brassica oleracea Unpublished (2001)

AUTHORS

TITLE

JOURNAL

COMMENT Other GSSs: BOHKJ17TF

Contact: Chris Town

TIGR 9712 Medical Center Drive, Rockville, MD 20850, USA.

Tel: 301-838-3523

Fax: 301-838-0208

Email: cdtown@tigr.org

DNA is from a doubled haploid provided by Tom Osborn.

Seq primer: TR

Class: sheared ends.

FEATURES

source 1..612

Location/Qualifiers

/organism="Brassica oleracea"

/strain="T01000DH3"

/db_xref="taxon:3712"

/clone="BOHKJ17"

/clone_11b="BOHK"

/note="Vector: PHOS1, Site_1: BstXI; 2-3 kb sheared genomic DNA inserted into PHOS1 using BstXI linkers"

BASE COUNT 198 a 113 c 94 g 207 t

ORIGIN

Query Match 4.3%; Score 93.2; DB 12; Length 612;

Best Local Similarity 67.4%; Pred. No. 0.0015;

Matches 225; Conservative 0; Mismatches 93; Indels 16; Gaps 6;

QY 1769 gttagaacaacagatacgagttgcctgaatgcglttaagagtcgtgactggtgctggg 1828

DB 19 GTTAGGGAAGAGGAGTTTAAATTAGTGTCTTCTATATACAGAGCCTGGAGATGGTTCCTTGGG 78

QY 1829 cagtaagcaatacaaaaagaacaaacccca--aaacccaagacagatataact--ccgata 1885

DB 79 CAGTAAGCAATTCACAAACAACAATACCCCCCAACCAACACAGATATACCCCAATA 138

QY 1886 ccaacacacaggtatcatctactattcaaaaaaacaacacaggtlaagtaagaagact 1945

DB 139 CGAATCAGAGGTTATATATACATTTA-TTAAACAATAGTATATATGATATAACT 197

QY 1946 -----ccttacaagattatatacttaactgaagctgacttaattagctctttagt 1997

DB 198 AAGCAATGCTCTACAGATTATATATCTGATC--TCTGGACTTTTACCTCTTATAT 254

QY 1998 accaatatagtcgcccacatctgtgcgtcatcacatcttcttcttccctaa 2057

DB 255 GGCATATATATACCAATTTTGTGTGCGC-CACCAAAATTTATGTGTTATTTTCCCA 313

QY 2058 tcaattagactcacaatactcttaaaaaagaatatt 2091

DB 314 TACACTACACTCCCATTTTATTAATAATCATTTT 347

RESULT 12

BH462017

LOCUS BH462017 605 bp DNA linear GSS 13-DEC-2001

DEFINITION BOHN243TR BOHN Brassica oleracea genomic clone BOHN243, DNA sequence.

ACCESSION BH462017

VERSION BH462017.1 GI:17652958

KEYWORDS GSS.

SOURCE Brassica oleracea.

ORGANISM Brassica oleracea

REFERENCE Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M. Whole genome shotgun sequencing of Brassica oleracea Unpublished (2001)

AUTHORS

TITLE

JOURNAL

COMMENT Other GSSs: BOHN243TF

Contact: Chris Town

TIGR 9712 Medical Center Drive, Rockville, MD 20850, USA.

Tel: 301-838-3523

Fax: 301-838-0208

Email: cdtown@tigr.org

DNA is from a doubled haploid provided by Tom Osborn.

Seq primer: TR

Class: sheared ends.

FEATURES

source 1..605

Location/Qualifiers

/organism="Brassica oleracea"

/strain="T01000DH3"

/db_xref="taxon:3712"

/clone="BOHN243"

/clone_11b="BOHN"

/note="Vector: PHOS1, Site_1: BstXI; 2-3 kb sheared genomic DNA inserted into PHOS1 using BstXI linkers"

BASE COUNT 190 a 156 c 116 g 143 t

ORIGIN

Query Match 3.8%; Score 80.8; DB 12; Length 605;

Best Local Similarity 68.8%; Pred. No. 0.07;

Matches 143; Conservative 0; Mismatches 57; Indels 8; Gaps 2;

QY 8 cagccactctatagctctctgactaagaacttcatttcaa---aatcgaatttcta 64

DB 314 CATCCACTCTCCAGCTCTACTACTACATCAACCAACCAAGAGATTCCATTTC 373

QY 65 attctagtttcaagcttcgt-----acggagaaaatgaagagatcgatgaagga 119

DB 374 AATTCGAATTTGAAATTTCATTCGAGAGAGTTTCGTGATGAAGTGAATGCGTGGAGGA 433

QY 120 attcaagcgcaagcctgcgcgacgcgctgcgcgggaagaagtcggtcgaacgcat 179

DB 434 ATTCAAGGCTTAAGGTGGGGGACGCCCTTGGCCGCAAGAACTTCGCTGATGAGATT 493

QY 180 taagccggaagaagacagatctctcc 207

DB 494 CGATCCTGAAGATAGCTGTGATCTCTCC 521

RESULT 13

LOCUS A2686821 791 bp DNA linear GSS 14-DEC-2000

DEFINITION ENPL174TR Entamoeba histolytica Sheared DNA Entamoeba histolytica genomic, DNA sequence.

ACCESSION A2686821

VERSION A2686821.1 GI:11823967

KEYWORDS GSS.

SOURCE Entamoeba histolytica.

ORGANISM Entamoeba histolytica

REFERENCE Loftus,B., Van Aken,S. and Fraser,C.

2

REFERENCE AUTHORS TITLE JOURNAL	2 (bases 1 to 132699) Ecker,J.R. Direct Submission Submitted (25-FEB-1999) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA 3 (bases 1 to 132699) Ecker,J.R. Direct Submission Submitted (07-MAY-1999) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
REFERENCE AUTHORS TITLE JOURNAL	4 (bases 1 to 132699) Cheuk,R., Shinn,P., Brooks,S., Buehler,E., Chao,Q., Johnson-Hopson,C., Khan,C., Kim,C., Altafi,H., Bel,B., Chin,C., Chou,J., Choi,E., Conn,L., Conway,A., Gonzalez,A., Hansen,C., Houng,B., Koo,T., Lam,B., Lee,J., Lenz,C., Li,J., Liu,A., Liu,J., Liu,S., Mukharshy,N., Nguyen,M., Palm,C., Pham,P., Sakano,H., Schwartz,J., Southwick,A., Thavert,A., Toriumi,M., Vaysberg,M., Yu,G., Davis,R., Federspiel,N., Theologis,A. and Ecker,J. Submitted (28-JUN-2000) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th and Hamilton Walk, Philadelphia, PA 19104-6018, USA On May 7, 1999, this sequence version replaced g1:4731042. Location/Qualifiers 1. 132699 /organism="Arabidopsis thaliana" /db_xref="taxon:3702" /chromosome="1" complement(join(144..278,360..491,576..673,814..1435, 1749..1867,2161..2310,2441..2565,2698..2742,2837..2902, 2979..3031,3117..3234,3314..3415,3498..3655,3741..3818, 3894..3985,4054..4104,4138..4420)) /note="similar to Na/H antiporter dbj BA089487.1" /codon_start=1 /evidence=not_experimental /product="F10B6.1" /protein_id="AAF79251.1" /translacion="MSTIGALNPKSPERKAIASSVSAENDSPVDVITAGTSLV GACRYLNGRVRVTVLVVIGFISGLICILIKLISVSISMKILLEYGTNHL KICLGRINRNGINPDILLAVFLPVLLFESSFSDVHRIKRCOMVLAAGVLIST PCIGALIKLTPINMDKTSLLDGLGATDPVAVVALLKEIGASKMTTLIDGSLM NDGVSVVFDLFEKYMGNHNSDMSIIRKLVNDSGAVGIGIAGIASVFWKLFND TVQITVLSASYFAVYQAQEMAGVSGILFMIIAGMFAAFARATFGKDSHSLHPN EMAYLIANTLVFMSGVIAESVLSGOTISYKVSOPILRPDEFILFLEDVYGN SMRFLFLYLVOLSRGVVGLPILCRSGYGDMSKSTILTWGSLRAVLOFLTG GIVPLTVVNGSTQDILLHLRMTLTATKRIILEYTFEEMAKTRAKAPENGDDEL GSADPTVIRHISLKDLEGRQVPHDYEAGSDPTNIMDIRIRFLNGLISOCVANS HCQTSVPKVAIDRLPESGVAAYWEMLDGRTTOCTANVMQSVDEADLVSTSS LSDMRGLEPRVHPNPKYFLOSKILPHLVTNHLVERLESACYISSAFRLHRIAROC LHLEFGNSINASTVINSEVEGEAKOFLPDVRSPOVLVSLKRTQVTHVYVNLNG YIKMLKVGGLGKEVSHLHVQVSDKLRLRHPSLKLPVNDLITNSPLKDRSSP RSALGCTDA"
CDS	join(5650..5801,6081..6121,6330..6819) /note="unknown protein; similar to EST emb F14394.1" /codon_start=1 /evidence=not_experimental /product="F10B6.2" /protein_id="AAF79216.1" /translacion="MRPPTTLLVGAIFSGAGVRSASDHRKKEGDVPLXANKY GPFNPSEYRYEDLPFCIEGVKKEKALEVLNGDRLVASPYKLNFRDKESEVYC NKRLSKEVKEOFKRAVKEDYFQMYDOLPIWFGVKGDKDIKSPSEKVELYKHO FELIYNKDVLEISARMDPSLNDLTDKEDVDAEMVTVKMKETETPEPKMEKYSMS SSLPNHEIFHMFSTINSCVYLLTGLATILMKYLNDEPK" join(6881..7029,7140..7296,7392..7876,7961..8180) /note="similar to transmembrane 9 superfamily member 2 g1:4758874; similar to dbj AAV42489.1, gb A1596021.1, emb F14395.1" /codon_start=1 /evidence=not_experimental
CDS	complement(13762..14781) /note="hypothetical protein" /codon_start=1 /evidence=not_experimental /product="F10B6.5" /protein_id="AAF79219.1" /translacion="MDDGFRMNGYEPAAATFKNGILQLMSTIDRNTKPLPGHDP NLMGNGSYNHQEPPTIHMSYVWVINOCKRPFMLVPTATPYGAVLEPSSAPSQ NMLHILQTEENDPKLEELIYVOTKRNARAKGSPKATKPKPKDEMSNNNNNT NTRVPRPAKSDVLYNGVSMDSIGLPVLCITGTAPOCCYRPGCGGSCACQTTIS MPLPSTRRGARISGRKMSGACAFKVLLEKLASDGFNGNPIDLKSHMRHSTNKFV TTR"
CDS	complement(17635..17815,17901..18019,18141..18194, 18300..18406,18547..18611,18742..18950)) /note="hypothetical protein" /codon_start=1 /evidence=not_experimental /product="F10B6.7" /protein_id="AAF79220.1" /db_xref="GI:8778211" /translacion="MSTCYRECMNRHNAKISYVAIDGCREYSPSTGLCVAGCH RSYHRIDVSSPOINHTREPTSLRVSOLARIKKTAEENEFEEDTETRETEPK MTVQRNRKKTFAEDREANKVDAAKIGMLTKKRALREIRIAPCEIGYTRHFKTW NNKKFYH"
CDS	complement(join(17635..17815,17901..18019,18141..18194, 18300..18406,18547..18611,18742..18950)) /note="hypothetical protein" /codon_start=1 /evidence=not_experimental /product="F10B6.8" /protein_id="AAF79249.1" /db_xref="GI:8778240" /translacion="MEFKAYFIWVROPRLQCEPIVILFCROMTCLKKORIFLFW VDEELHLODPLPKVYVFLIRIRMPDKOVKQSSSTKQDSLVSVCABALGNE IVRPYEVKSSFKSLSPKSGSKVTFQALKVGSLTFVVEAANDKAGNMTQSK EOSTLIVSESSFDEVMNNKSKSDELCTVLAKIFGSIAVAIAVYVGVARVLLPFV

CDS
 MKKDLGSSSSONGSSDSEDJDSIDF
 complement(join(20071..20409,20499..20573,20648..20721,
 20800..20887,20957..21108,21188..21356,21448..21906,
 22006..22275,22357..22570,22840..22917))
 /note="hypothetical protein"
 /codon_start=1
 /evidence=not_experimental
 /product="F10B6.9"
 /protein_id="AA079248.1"
 /db_xref="GI:8778239"
 /translation="MEIESPTSLCFRTNTTNCALNRELQKIVWDIGESDAEKDRMLM
 ELEKECEIYRKRDEANASKAOLHOSVLEAEIALALGVFNHSPMKKEGSK
 SLKEKLAIVPMLEDRLOKDERMKOFVDIKAOLEKSGESGSDQNTMGSLAL
 DEODTLRKLINEYQHLRSLOKESDRLNLDVNVNHLICVGLGVDFQGVSEVP
 SLARHDSQSTNISDPTLDGLHMHILKTRRSYRQVKNACNAILLVYTDYHEPA
 LAEPISLVQSLDVAGSLPELNMMDTSQEBRTKASVSYVRSSSDITEPNILISST
 IEQVSAEVDENKELKSRMELVMKRTLELNCRAHIEDTSTSLKSTALDSGT
 KRIYDSEILTNIELHINKIEEASHKEIIDRIDMLACSEENMLEEYNLDSTBY
 AGRGHVNLKAEARITVNRKIPSMVDNLIKTKLMEDETRSPFLYGVRLSLEDY
 KLTNRKOEERKRRYRDKMODLIRKRESYSGSPRSNSVRYKTNNGDASVPR
 TPRNSGATNMIDMTTPRSTSSHONRYEVRKSTAPLNPAITREDSYSTYSV
 CGSEPDSPLYN"
 join(26661..26948,27454..27610,27713..27814,27911..28038,
 28342..28470,28565..28632,28747..28969)
 /note="similar to acid phosphatase type 5 emb|CAB6398.1;
 similar to Est pblN96048.1"
 /codon_start=1
 /evidence=not_experimental
 /product="F10B6.10"
 /protein_id="AA079221.1"
 /db_xref="GI:8778212"
 /translation="MTYVYRDKITTKSTPIELFELFCEFNLSMATLKHKPVNLVY
 VYNLIIFFSSSHSTAEELRLIQLPSKTDTSVFLVIGDMGRGSTNQGVALQMEIGE

Query Match 52.1%; Score 1117.4; DB 8; Length 132699;
 Best Local Similarity 99.9%; Pred. No. 2.6e-258;
 Matches 1118; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 actgcatcagccactctctagctctgactaagaaattccatttcaaaattcgatt 60
 Db 50501 ACTGCATCAGCCACTCTCTAGCTCTGACTAAGAAATTCATTTCGAATT 50442
 QY 61 tctaatctttagtttcaagctctctgtagcgaagaaatgaagagatcgcatgagaa 120
 Db 50441 TCTAATTTCTAGTTTCAACCTTCGACGAGAAATGAAGAGATCGCATAGGAA 50382
 QY 121 ttcaaaagcgaagcctgtagcgaagcgcgtctgcccgggaagaaattccggtcgacgcgatt 180
 Db 50381 TTCAAAAGCGAAGCCTGACGACGCGCTGCCGGGAAGAACTCCGCTCGACGCGATT 50322
 QY 181 aagcgggaagaaagacaaatctctccgcttcttcttcaatacctctcttgagaacaa 240
 Db 50321 AGCGCGGAAGAGACAAATCTCTCCGCTTCTTGTTCATATCACTCTTGAGCAAA 50262
 QY 241 aatcgagatcctctgctctctgctgactctctgctcgaatttgtagctgtagacaaagt 300
 Db 50261 AATCGAGATCTCTGCTCTCTGCTGATTCCTGCTCGATTGCTGATGACAAAGT 50202
 QY 301 ttctctggttcgagcagatcgagaagcctcgaaatccgaagaagaaaccttaattgaaga 360
 Db 50201 TTCTCTGTTGTCAGACGAGTCCGAAAGAGCTCGAATTCGAAACACTCTAATTGA 50142
 QY 361 ggtagaagattcttaaacctggttataatgtgaagagagacattggtgattcgaaattcg 420
 Db 50141 GGTAGAAGATTCTTAACCTGGTTATATGTGAAGAGACATTGGTGCATTCGAAATTCG 50082
 QY 421 aaggaattacgaagctctactactaagctacacaaagaaagaaagaaatgaatgaatgaat 480
 Db 50081 AAGGATTACGAGCTTACTCTAAGCTACACAAAGAAAGAGAGAGATGAGATCGAAGT 50022
 QY 481 aagcgaatcgtctctggtgattcgaaattctgctgctgtaataagaaattgaattga 540
 Db 50021 AAGCGAATCGTCTTGTGTTGATTGAAATTCGTGCTGGAATTAAGAGATTGAATGTA 49962

QY 541 ggggaataaataaagacacaaatgagatctcttctcaagatccgatgtgacctcgc 600
 Db 49961 GGGGAATAAATAAATTAACACAAATGAGATCTCTTCTCACGATCCGATTCACCTTCCG 49902
 QY 601 cgaacagctctccacaaagcgaagtttgaaatttcgaatcgaagaataaagaagcgaagt 660
 Db 49901 CGACAGATGTCCTCCACAGCCGGAGTTTGAAATTCGATTCGAGAAATTAAGAGAGCCAGT 49842
 QY 661 cgttctgctatctcgtgagttgagttcgttccaaatttcgggagcgttcaggagagc 720
 Db 49841 CGTTTCTGCTATATCTGAGTGTGAGTCTGCTCCAAATTCGGGAGCGTTACCGAGAGC 49782
 QY 721 tgaataagaaatltgaatctccaaagccgaagcgttcgtgaaagctgattctctc 780
 Db 49781 TGATACGAAGAAATTTGAATCTCCAAAGCCGAGAGCTTCGTGAACCTGATTCCTCT 49722
 QY 781 tgaatcgcgaagaaatltgaagccgaagcttgagataagtcgagtcgtctgacgc 840
 Db 49721 TGATTCGGCCCAAGGAATTTGAAGCCGAGCTTGAGATAGTCGATGCGTCTGATCTCGC 49662
 QY 841 ttgctctgagaatcttcgcgaagaggttcgagttctctcgaatgaatgaatcagaa 900
 Db 49661 TTGCTCTGAGAAATTTCTCGAAGAGTTTGGATTCTCTGATGATGATCATCTGAGCA 49602
 QY 901 acgttcaagatatattacacagattccgaacttcgattactcgaattacactccgtccat 960
 Db 49601 ACGTTCAAGATATATTCACAGTATTCGAGTTGATTCGATTCGATTCGATTCGATTCGAT 49542
 QY 961 ctctctcgaactctgacgaagaaatctctctgagaatctctctgattctcctattaca 1020
 Db 49541 CTCTTCGACACTCTGACGAGAAATTTCTCGAAGAAATTTCTCTCTGATTCCTATTTTACA 49482
 QY 1021 ttctcgtctctgtaacctccagttcgaagaacaaattctgtagatccagattcccaaga 1080
 Db 49481 TTCTCGCTCTGTAACCTCCAGTTCAAGAAAGTTCTGTAATCCAGATTCCTCCAAAGA 49422
 QY 1081 ttctgacatctctgacgaagaagaatcactctgaaatt 1119
 Db 49421 TTTTGATCTTCTTCTCGAGGAAGAAATTCACCTCGAAGT 49383

RESULT 2
 AX345803
 LOCUS AX345803 5204 bp DNA linear PAT 01-FEB-2002
 DEFINITION Sequence 874 from Patent WO0200928.
 ACCESSION AX345803
 VERSION AX345803.1 GI:18493689
 KEYWORDS
 SOURCE
 ORGANISM
 synthetic construct.
 synthetic construct
 artificial sequence.
 1 (sites)
 REFERENCE
 Olek,A., Piepenbrock,C. and Berlin,K.
 TITLE Diagnosis of diseases associated with the immune system
 JOURNAL Patent: WO 0200928-A 874 03-JAN-2002;
 Epigenomics AG (DE)
 FEATURES
 Location/Qualifiers
 1..5204
 /organism="synthetic construct"
 /db_xref="taxon:32630"
 /note="Chemically treated genomic DNA (Homo sapiens)"
 BASE COUNT 1463 a 72 c 1014 g 2655 t
 ORIGIN

Query Match 3.7%; Score 78.4; DB 6; Length 5204;
 Best Local Similarity 53.2%; Pred. No. 4.1e-08;
 Matches 166; Conservative 0; Mismatches 146; Indels 0; Gaps 0;

QY 1833 aagcaatcaaaaagaacaaacccctaaacacgaacagatatactcctgataccaaacac 1892
 Db 2507 AAAAAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 2448

Query Match	3.7%	Score 78.4	DB 6	Length 349980
Best Local Similarity	53.2%	Pred. No. 5.2e-08		
Matches 166	Conservative 0	Mismatches 146	Indels 0	Gaps 0
QY 1833	aagcaatcaaaaagaacaaacaccttaaacacgagacaglataccgcgataccacac	1892		
Db 60597	AAAAAAAAAAAAAAAAAATCTAATAAATAATAAAAAACAAAAAATAAAAAAAC	60538		
QY 1893	acaggtatcatctactattacaanaacaacaagtaagtataagaactcctctac	1952		
Db 60537	AAAAAAAAACAAAAAACAAAAAATACTTAACACATATAAAAAACCTATATATAA	60478		

Query Match	3.3%	Score 70.6	DB 5	Length 4435
Best Local Similarity	78.0%	Pred. No. 3e-06		
Matches	85	Conservative	0	Mismatches 24
				Indels 0
				Gaps 0

Query Match	3.1%;	Score 67.4;	DB 3;	Length 6372;
Best Local Similarity	51.1%;	Pred. No. 1.8e-05;		
Matches 158; Conservative	0;	Mismatches 151;	Indels 0;	Gaps 0;

RESULT	7		
AX344558/c			
LOCUS			
AX344558	349980 bp	DNA	1 linear
Sequence 9 from Patent WO0200932.			PAT 01-FEB-2002

BASE COUNT	97534 a	3966 c	76586 g	171894 t
ORIGIN				
Query Match	3.1%	Score 67.2:	DB 6:	Length 349980:
Best Local Similarity	51.3%	Pred. No. 2.6e-05:		
Matches 156:	Conservative 0:	Mismatches 146:	Indels 0:	Gaps 0:
QY 1841	aaagaacaaaaccttaaacacgagacacagtaactcgcataccacacacagta 1900			
Db 69065	AAAATAACGTAACCCGAAACGAAACGAACTTACATTAACCTAAATACACGACATACACT 69006			
QY 1901	tcattcatttcagaaacaaacacacgaagtaagtaatacgaacccctctaaagttcat 1960			
Db 69005	CCAACTTAATTAACAAATTAATAACCTCCACCTTAATAAAATATACAAATAATTAAACC 68946			
QY 1961	atacttaatcgagcgcgaacttaatagctctctagatataccaattatagtcgcacatlt 2020			
Db 68945	AAACGTATATACGAACGGCTTCAATCCACACTATATTAATAAAACCTAAACAAATAATGCG 68886			
QY 2021	gtgcgcgcacatacattatlttctatttccctaatcattcattagactctcatatltcta 2080			
Db 68885	TTTAAACCTTAATAAAACAAATAATTAATTAACATTAACCGAATATACCTACATCTCAACCTTAA 68826			
QY 2081	aaaagaatattccctgttgcgaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2140			
Db 68825	ATAACAAATCAAAATTCATCTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 68786			
QY 2141	aaaa 2144			
Db 68765	AAAA 68762			
RESULT 8				
AFJ10889/c	7736 bp	DNA	linear	INV 06-MAR-2001
LOCUS				
DEFINITION	Dictyostelium discoideum Racd (racd) gene, complete cds; and unknown genes.			
ACCESSION	AFJ10889			
VERSION	AFJ10889.1	GI:12007300		
KEYWORDS				
SOURCE	Dictyostelium discoideum.			
ORGANISM	Dictyostelium discoideum.			
REFERENCE	Eukaryota; Mycetozoa; Dictyostelidia; Dictyostelium.			
AUTHORS	1 (bases 1 to 7736)			
TITLE	Rivero,F., Dislich,H., Glockner,G. and Noegel,A.A.			
JOURNAL	The Dictyostelium discoideum family of rho-related proteins			
MEDLINE	Nucleic Acids Res. 29 (5), 1068-1079 (2001)			
REFERENCE	2 (bases 1 to 7736)			
AUTHORS	Rivero,F., Dislich,H. and Noegel,A.A.			
TITLE	Direct Submission			
JOURNAL	Submitted (02-OCT-2000) Institut fuer Biochemie I, Medizinische			
FEATURES	Fakultaeet, Universitaet zu Koeln, Joseph-Steitzmann-St. 52, Cologne			
SOURCE	50931, Germany			
	location/Qualifiers			
	1..7736			


```

/organism="Dietystelium discoideum"
/strain="AX4"
/db_xref="taxon:44689"
join(<856..1007,1160..>4670)
/product="unknown"
join(856..1007,1160..4670)
/note="5' region identified manually, therefore unsure;
similarity to mouse SH3-domain containing adaptor protein
(central part) and rat L-glutamine aminohydrolase
(C-terminal part)"
/codon_start=1
/product="unknown"
/db_xref="GI:12007301"
/translation="MFKHOSNNLYVYFNPINPVANVTKODIHAETSSIGGVKGL
MCEYCFIKSKMAIKDDYQVQMDKOQCYKMHLLSETESAKVAVLEVBATE
EELSLPHQVIDITISKDSGMEGSKOONMGVFPNSRVDPEPTDQEPEDDSAEAT
KEEISKPIQSVSLGINISLGSMDLKNLKVAKRTVTSADIEAHNNMNASIES
NSSSSSSSSSSSSSGGFVAVKPSGAVTQORQKVLNNROSTILSGPAIOSIP
SSNTTTTNTTINAVOPLIQKPKONTLEINTGNSIPATTSKESQSTKESSSGDN
KITSLEKPKLKITAISALGSHSSNESSGGGNNNGSKTKAVVVDPSKEPNEIN
LKRDITIVLAKDSSGMGGINOSTQATCMFSNTEVEIKKEKLPSPSSSENQPK
IVKSPKSSSSSQPQVEAKLSLSDTPKLETKAKGAAKGRKPPSVRASTILSD
EKAREKLKOSFYSHVVSOKELSMYNNEDGHLDDQPTSPS1SRSGSFTTP
QOOOOSVDKIERPSKPPKRTAANSSLNINNEPPSKNPPPKSLKPLSTD
GYEPSPSSSTPTTSGEDTQKSPTLQRTKVAAPPOQNEGLPSLVGKPPPK
PTPKRAIOTNESTAPAPPSHEVNKANKPAKPSPPIGNSQELIPTDSSPPNAT
TPQTPPKRKVAKPPVQOQPSNMENTITQFKLKSPPASPTSPSSPTIDPTTK
PPAPPLKSAANAAPPSPNEATTSPPLKPPKALVAGNSAPSTBPSQPOOP
LPSTHKRNLSNPTTNNKPPPPKPKQLKPLKPPNDICEVSNCSNETIERSDMS
QETTTTTHNPSLIDSFYAVCSNDGFWGNDNDHCVPMEQVITYPLYSILCE
BIGDEVSKFISDAPRDSYDNOKPLNDOKKSHNPTLGAOLITANLSVNIQMHLE
STORNIIOELVNCSSVCDMWYLSQSESDSLGLKSSGLIKOPDVLDPFY
QINSIOINSKOVSVLAATLANDGCPSPQDLKAPKNTISCTIDILICNDPTTSSI
FDNNIDINNNKTIPIISDSGYLMIITPLGLMGVSIISPCNNNNNNNNNNNDNFHYEN
NMHLEFCRKLCSILDQ"
join(5116..5180,5293..5333,5523..5590,5805..5914,
6023..6634)
/gene="racd"
/note="corresponds to EST clones SSB614, SSG550, S8J204
and SSK792 from the Japanese cDNA sequencing project; cDNA
(partial) already described in L11594."
/product="Racp"
5116..6634
/gene="racd"
join(5163..5180,5293..5333,5523..5590,5805..5914,
6023..6550)
/gene="racd"
/note="Rho-GTPase; lacks prenylation motif characteristic
of Rho proteins"
/codon_start=1
/product="Racp"
/protein_id="AAG45121.1"
/db_xref="GI:12007302"
/translation="MATGIKRTKVVVVVGDGAVGKTSLLIITYTKAPKRDVYPTFDN
FNCLEMYNKNPVNLVMTAGQEDYDLNPOTDVFILICYSVKADSDINIKYK
LPEINOTNOGPTIIVGTKTDLREDKTLTLOESKOEPVDEGVALLAKIGAVOFF
ECSATGNGVDITFAAATKAAPNPAVTSPTSSKSGSPSSSTSSKPKTTTTTSS
SSSSPPASTAKPAGCKLMSGLPRKKDKDKRAK"
6612..6617
/gene="racd"
join(<7278..7533,7645..>7736)
/product="unknown"
join(7278..7533,7645..>7736)
/codon_start=1
/evidence-not_experimental
/product="unknown"
/protein_id="AAG45122.1"
/db_xref="GI:12007303"
/translation="MSIOSDEDFEELCKDEYLVKPPETIKIDDIHQYFANPFLI
EYKOLISPTKNSYVAVQYIRGLIKCVTSFWSLTPNOKNDKSNITWLTIESVPLEQF
ALSLVFKLKSRVLA"
BASE COUNT 3205 a 1097 c 837 g 2597 t

```

```

ORIGIN
Query Match 3.1%; Score 65.6; DB 3; Length 7736;
Best Local Similarity 51.0%; Pred. No. 5e-05;
Matches 155; Conservative 0; Mismatches 149; Indels 0; Gaps 0;

QY 1841 aaaaagacaacaaaccctaaacaggaacagatctctccgatlacacacacagttla 1900
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 5348 AAAAAAAAAATACATCCACAGCACACCTACACTCACTACTACTTAAACAGTTTGT 5289
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1901 tcatctatttacaacaaacacaaaggaaglaaglaagaaactcctctacagattat 1960
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 5288 TTTTAAAAAATTTTAAAAAATTTTAAAAAATTTTAAAAAATTTTAAAAAATGTTAA 5229
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1961 atacttaatcagctgactaactaactagctctagatatacaattatgaagccacatt 2020
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 5228 TAAATCAAAACCTTGTGGTAAATTAATTTTGGATTAATTAATTCATTACCTTAATACCTGT 5169
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2021 gtgcgcacacacacattatcttattccctaattcattagactcctacattctta 2080
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 5168 AGCATATATATTTATTTATTTTCTTTCTTTTATTTGAAAAAATAATGATGT 5109
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2081 aaaaagatatctcgttgtaaaaaaataaaaaaataaaaaaataaaaaaata 2140
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 5108 TAAAAAATTTTAAAAAATTTTAAAAAATTTTAAAAAATTTTAAAAAATTTTAAAAA 5049
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2141 aaaa 2144
|||||
DB 5048 AAAA 5045

RESULT 9
AX344553/c 349980 bp DNA linear PAT 01-FEB-2002
LOCUS AX344553 Sequence 4 from Patent WO0200932.
DEFINITION AX344553
ACCESSION AX344553
VERSION AX344553.1 GI:18492439
KEYWORDS
SOURCE
ORGANISM
synthetic construct.
artificial sequence.
REFERENCE
1 (sites)
AUTHORS Olek, A., Piepenbrock, C. and Berlin, K.
TITLE Diagnosis of known genetic parameters within the mhc
JOURNML Patent: WO 0200932-A 4 03-JAN-2002;
Epigenomics AG (DE)
FEATURES
source
1..349980
Location/Qualifiers
1..349980
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo
sapiens)-original length of seq 1: 3.673778 <223>-split as
follows.-seq 01 0.000.001 TO 0.349.980-seq 02 0.300.001
649.980-seq 03 600.001 949.980-seq 04 900.001
1.249.980-seq 05 1.200.001 1.549.980-seq 06 1.500.001
1.849.980-seq 07 1.800.001 2.149.980-seq 08 2.100.001
2.449.980-seq 09 2.400.001 2.749.980-seq 10 2.700.001
3.049.980-seq 11 3.000.001 3.349.980-seq 12 3.300.001
3.649.980-seq 13 3.600.001 3.673.778 <223>-Original length
of seq 2: 3.673778 <223>-split as follows.-seq 14
0.000.001 TO 0.349.980-seq 15 0.300.001 649.980-seq 16
600.001 949.980-seq 17 900.001 1.249.980-seq 18 1.200.001
1.549.980-seq 19 1.500.001 1.849.980-seq 20 1.800.001
2.149.980-seq 21 2.100.001 2.449.980-seq 22 2.400.001
2.749.980-seq 23 2.700.001 3.049.980-seq 24 3.000.001
3.349.980-seq 25 3.300.001 3.649.980-seq 26 3.600.001
3.673.778"
BASE COUNT 86882 a 5859 c 85073 g 172166 t
ORIGIN
Query Match 3.1%; Score 65.6; DB 6; Length 349980;

```

OY		1857	ctaaaccaggacacagatatactccgaattaccaacaacagaagttaactcattactaatlttaca	1916
	I I I I	I I I I	I I I I I I I I I I I I I I I I I I	
	I I I I	I I I I	I I I I I I I I I I I I I I I I I I	
Dd	49199	CAATTACACAACAGATTTTACTTAAMCCGAACCTAAATTATAATATTCGAAAACAAAAA		49140
OY	1917	aaacnaaacacaggttaagtaaagaactccctctacgagtttatatacttaacttgtagctg		1976

QY 2013 caccatttggf---cgctacatacacatllatttcttattlltcctaatlcttagactc 2065

Db 251631 AAAAATATCTACACGATATCTCAAAAAATTCCTTTCAAAACAAAAAACTTCAATAC 251572
QY 2070 tcatattctaaagaataattccctgttgtaaaaaaagaaaaaagaaaaa 2129
Db 251571 AATATCACTAAACAAAAATTTATCTATCTCTTAAAAAATAAACCAAAATATTT 251512
QY 2130 aaaaaaagaaaaa 2142
Db 251511 AAACAAAAA 251499
RESULT 12
LOCUS AX185705/c
DEFINITION Sequence 1400 from Patent WO0142467.
ACCESSION AX185705
VERSION AX185705.1 GI:15137111
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 545)
AUTHORS Schlegel,R., Deeds,J., Berger,A. and Zhao,X.
TITLE Genes, compositions, kits, and methods for identification,
assessment, prevention, and therapy of cervical cancer
JOURNAL Patent: WO 0142467-A 1400 14-JUN-2001;
Milleium Predictive Medicine, Inc. (US)
FEATURES
Source location/Qualifiers
1..545
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 200 a 40 c 23 g 209 t 73 others
ORIGIN
Query Match 3.0%; Score 63.4; DB 6; Length 545;
Best Local Similarity 54.5%; Pred. No. 0.00015;
Matches 91; Conservative 0; Mismatches 76; Indels 0; Gaps 0;
QY 1978 actaatagccttagatataccattatagccaccattgctgcctacacat 2037
Db 280 ANNTTTTCCCTNNANNTTTTNNANNTTTTGTGNNATTTTCCCTCCCAANTTTT 221
QY 2038 ttattcttatttcccaatcatcagctctcatattcttaagaagataattccttg 2097
Db 220 TTTTNNNNNTTTTNTTNTTNTTNNANNTTTTNNAAAAAANTTTTNTTAAANT 161
QY 2098 ttgaaaaaagaaaaaagaaaaaagaaaaaagaaaaa 2144
Db 160 TTTTNNAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 114
RESULT 13
LOCUS AY069454
DEFINITION Drosophila melanogaster LD17744 full length cDNA.
ACCESSION AY069454
VERSION AY069454.1 GI:17862243
KEYWORDS FLI.CDNA.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1090)
AUTHORS Stapleton,M., Broksstein,P., Hong,L., Aghayani,A., Carlson,J.,
Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E., George,R.,
Gonzalez,M., Guarini,H., Li,P., Liao,G., Miranda,A., Mungall,C.J.,
Nunoo,J., Pacleb,J., Paragas,V., Park,S., Phouanavong,S., Wan,K.,
Yu,C., Lewis,S.E., Rubin,G.M. and Celinker,S.
TITLE Direct Submission

JOURNAL Submitted (10-DEC-2001) Berkeley Drosophila Genome Project,
Lawrence Berkeley National Laboratory, One Cyclotron Road,
Berkeley, CA 94720, USA
COMMENT Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory
Berkeley, CA 94720
This clone was sequenced as part of a high-throughput process to
sequence clones from Drosophila Gene Collection 1 (Rubin et al.,
Science 2000). The sequence has been subjected to integrity checks
for sequence accuracy, presence of a polyA tail and contiguity
within 100 kb in the genome. Thus we believe the sequence to
reflect accurately this particular cDNA clone. However, there are
artifacts associated with the generation of cDNA clones that may
have not been detected in our initial analyses such as internal
priming, priming from contaminating genomic DNA, retained introns
due to reverse transcription of unspliced precursor RNAs, and
reverse transcriptase errors that result in single base changes.
For further information about this sequence, including its location
and relationship to other sequences, please visit our Web site
(http://fruitfly.berkeley.edu) or send email to
cdna@fruitfly.berkeley.edu.
location/Qualifiers
1..1090
/organism="Drosophila melanogaster"
/strain="y; cn bw sp"
/db_xref="taxon:7227"
/map="99D1-99D1"
1..1090
/gene="CG7866"
/note="alignment with genomic scaffold AE003771"
/db_xref="FlyBase:FBgn0039725"
39..500
/gene="CG7866"
/note="Longest ORF"
/codon_start=1
/db_xref="FlyBase:FBgn0039725"
/product="LD17744p"
/protein_id="AAL39599.1"
/db_xref="GI:17862244"
/translation="MSLGNDPWHLKNITVYKIGISGSANNSFEDNFTPLPGNSN
YKDGKIEPLPSDDYLKLERLARVOKNKLIDNRDRDCMRGLISSECVPISI
FEQFLIDAPIESGRHLRHLLPVOAVNGEFTFIVHDALQSAEEAQERE"
BASE COUNT 341 a 244 c 266 g 239 t
ORIGIN
Query Match 3.0%; Score 63.4; DB 3; Length 1090;
Best Local Similarity 66.4%; Pred. No. 0.00015;
Matches 91; Conservative 0; Mismatches 46; Indels 0; Gaps 0;
QY 2008 agtgcacacattgctgcgtcacaacattattcttattccctaatcattgac 2067
Db 939 AGAGTAACCTGCTTTATAGCTTATGATCATGTCGACCCCTATTATTATC 998
QY 2068 ttcataattctaaagaagatttccctgttgtaaaaaaagaaaaaagaaaaa 2127
Db 999 TCAACACGATATTTTAACAAAGTATTTAAGTCACAAAAAATAAAAAA 1058
QY 2128 aaaaaaagaaaaa 2144
Db 1059 AAAAAAATAAAAAA 1075
RESULT 14
LOCUS AX188346
DEFINITION Sequence 4041 from Patent WO0142467.
ACCESSION AX188346
VERSION AX188346.1 GI:15139819
KEYWORDS human.
SOURCE Homo sapiens

QY 2030 atccacattatctcatttcccctaattcatlagaacttcataatctttaaaagaaata 2085
||| ||||||| || | ||| || ||| | ||| ||| |

```

1  COMPUTER READABLE FORM:
2  MEDIUM TYPE: Floppy disk
3  COMPUTER: IBM PC compatible
4  OPERATING SYSTEM: PC-DOS/MS-DOS
5  SOFTWARE: Patent in Release #1.0, Version #1.30
6  CURRENT APPLICATION DATA:
7  APPLICATION NUMBER: US/06/722,126A
8  FILING DATE: 08-OCT-1996

```

? FILING DATE: 06-APR-1995
 ? PRIOR APPLICATION DATA:
 ? APPLICATION NUMBER: IL 109257
 ? FILING DATE: 08-APR-1994
 ? ATTORNEY/AGENT INFORMATION:
 ? NAME: BROWDY, Roger L.
 ? REGISTRATION NUMBER: 25,618

TELEPHONE: (617) 227-74
TELEFAX: (617) 227-5941

```

; MOLECULE TYPE: CDNA
;
FEATURE:
;

```

Query Match	2.7%	Score	57.4	DB	1	Length	1172
Best local	Similarity	68.7%	Pred.	NC	2e-05		
Matches	79	Conservative	0	Mismatches	36	Indels	0
						Gaps	0

Oy	2090	tttccttglttgaaaaa	2144
Db	1079	TAAATTCATTCACAAAAATTAAAAA	1133


```

: CURRENT FILING DATE: 1997-06-05
: NUMBER OF SEQ ID NOS: 18
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 8
: LENGTH: 1215
: TYPE: DNA
: ORGANISM: Human
US-09-092-770-8

Query Match      2.7%; Score 57.4; DB 2; Length 1215;
Best Local Similarity 30.6%; Pred. No. 2e-05;
Matches 112; Conservative 68; Mismatches 180; Indels 6; Gaps 1

QY 1233 ataccgtctcatccctcgtctacgctccatcatgttcaatagatgtgaaagcaatggt 1292
DB 401 atwswgayytungarcnccarcatgmgwnsnathytungaytgytunytungarcttgyg 460
QY 1293 ctgacatggggcttcagcagaagacatgttctctaggaattgtgctgttgatcgatcc 1352
DB 461 argtntayacynltuagmgngaracnctlytarytngcnaragaylttlytgymntcyg 520
QY 1353 tgaagcaaggatcatctcaaaagcgaagaagactctaatactagtcggagatcgagtccta 1412
DB 521 tgytnacncararayaathaaayaarayatlytncarytnathgnaethacnwsnynt 580
QY 1413 cctctggccaccgaagattaaagaaatcaacccttacaacgacatccggaagaagacttca 1472
DB 581 tyathgcnwsnaarytngargarathtataygcncnaa-----ryuncargatlygent 634
QY 1473 ccatcagaaccctaagatatagccgycatgaaigtgltggcaatggagtgcgtgtcaag 1532
DB 635 aytlnacngaygngcngtgywngargargayathytmmgnaatggarytnathaytna 694
QY 1533 aagtcctaactcaaatagcttcacaccacacatcttcaactcttgygtctactta 1592
DB 695 argcnytnaarvtggargytnltgycngtnacnathlmsnltgytnaaytnlttytnc 754
QY 1593 aagctg 1598
DB 755 argtng 760

RESULT 11
US-09-222-851-8
: Sequence 8, Application US/09222851
: Patent No. 6165753
: GENERAL INFORMATION:
: APPLICANT: Coats, Steven R.
: APPLICANT: Bass, Michael B.
: TITLE OF INVENTION: No. 6165753el Cyclin E Genes and Proteins
: FILE REFERENCE: A-524
: CURRENT APPLICATION NUMBER: US/09/722,851
: CURRENT FILING DATE: 1998-12-30
: EARLIER FILING DATE: 1998-06-05
: NUMBER OF SEQ ID NOS: 18
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 8
: LENGTH: 1215
: TYPE: DNA
: ORGANISM: Human
US-09-222-851-8

Query Match      2.7%; Score 57.4; DB 4; Length 1215;
Best Local Similarity 30.6%; Pred. No. 2e-05;
Matches 112; Conservative 68; Mismatches 180; Indels 6; Gaps 1;

QY 1233 atactggtcatccctcgtctacgctccatcatgttcaatagatgtgaaagcaatggt 1292
DB 401 atwswgayytungarcnccarcatgmgwnsnathytungaytgytunytungarcttgyg 460

```

```
QY 1293 ctgacatggggcttcagcaagagacattgtttctagagatgtgctgtgtgacatcc 1352
      :| | : ||| | | :| | :| | | | | :| | | | | | | | |
Db 461 argltatyacnltncaymgngaracnlttlaytngcncargatlttlygmgnttla 520
      :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
QY 1353 tggagcaaaagatctcaaaagcgaagagactccatctactctgaggttcgaactta 1412
      || | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
Db 521 tgytnencaraargatthaayaraaayaglyncarytattngnathcnwnynt 580
      :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
QY 1413 ctctgscaccagaatgtgaagaaatacaaccttaacagcatccggaagaagacttca 1472
      :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
Db 581 tythbgcnwnaarytngargarathcaygcncaa-----tylncargarttygnt 634
      :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
QY 1473 ccattcgaacctgaatatagccgcgatgaagggtgagcatgagtggttcaag 1532
      :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
Db 635 aytgnacngaygngcnlywngargargayathymnagnatggaarytlnathathyna 694
      :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
QY 1533 aagtcctcaactcaaatgtcttcacaccacatctcaactctctgtgttactta 1592
      :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
Db 695 argcnytnaartgsgartylntgycngtnacnathatwnstgyltnaaylnttylnc 754
      :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
QY 1593 aagctg 1598
      :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
Db 755 argtng 760
      :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
```

RESULT 12

```
US-09-342-681C-12
; Sequence 12, Application US/09342681C
; Patent No. 6355782
; GENERAL INFORMATION:
; APPLICANT: Zonana et al.
; TITLE OF INVENTION: Hypohydrotic ectodermal dysplasia genes and proteins
; FILE REFERENCE: 52978
; CURRENT APPLICATION NUMBER: US/09/342,681C
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: 60/092,279
; PRIOR FILING DATE: 1998-07-09
; PRIOR APPLICATION NUMBER: 60/112,366
; PRIOR FILING DATE: 1998-12-15
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 12
; LENGTH: 3720
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (260)..(1606)
; NAME/KEY: misc_feature
; LOCATION: (2961)..(3673)
; OTHER INFORMATION: n represents a, c, t, or g
US-09-342-681C-12
```

```
Query Match 2.7%; Score 57; DB 4; Length 3720;
Best Local Similarity 70.8%; Pred. No. 4.3e-05;
Matches 75; Conservative 0; Mismatches 31; Indels 0; Gaps 0;
QY 2037 ttattcttatttccctaatactgaactcattcttaaaagaattcttc 2096
      ||| | | | | | | | | | | | | | | | | | | | | | | | |
Db 3615 tatataattgtctatgttaataattattattataagcctaataacattgatna 3674
      :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
QY 2097 gttcgaagaaaaaagaaaaaagaaaaaagaaaaaagaaaaa 2142
      | | | | | | | | | | | | | | | | | | | | | | | | |
Db 3675 aaagggaagaaaaaagaaaaaagaaaaaagaaaaaagaaaaa 3720
      :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
```

```
RESULT 13
US-09-111-730-5
; Sequence 5, Application US/09111730
; Patent No. 6274359
; GENERAL INFORMATION:
```

```
; APPLICANT: Hideharu Anazawa
; APPLICANT: Hiroko Shimada
; APPLICANT: Seiji Sugimoto
; APPLICANT: Toshimasa Sudo
; APPLICANT: Tatsuo Sudo
; APPLICANT: Yuzuru Ishimura
; APPLICANT: Matsuhiko Hayashi
; APPLICANT: Toshiaki Monkawa
; APPLICANT: Tadaaki Yoshida
; APPLICANT: Shu Wakino
; APPLICANT: Takao Saruta
; APPLICANT: Hiromichi Suzuki
; TITLE OF INVENTION: 25-HYDROXYVITAMIN D3-1a-HYDROXYLASE AND DNA ENCODING THE HYDRO
; FILE REFERENCE: 1074
; CURRENT APPLICATION NUMBER: US/09/111,730
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 5
; LENGTH: 2469
; TYPE: DNA
; ORGANISM: Rat
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (24)..(1526)
US-09-111-730-5
```

```
Query Match 2.6%; Score 56.6; DB 4; Length 2469;
Best Local Similarity 60.9%; Pred. No. 4.5e-05;
Matches 92; Conservative 0; Mismatches 59; Indels 0; Gaps 0;
```

```
QY 1993 agtatccattatagtcgacacattgctgcgcctacacattattcttatttc 2052
      || | | | | | | | | | | | | | | | | | | | | | | | |
Db 2319 agtgcacactactactcttcacaccactcttcgcacccagcctgtctttatat 2378
      :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
QY 2053 cctaattcattagactcctcattcttaaaagaataattcctgttgaagaaaaa 2112
      || | | | | | | | | | | | | | | | | | | | | | | |
Db 2379 ttaaaaattgtattattattatttccaataaattgttacccttgaaagaaaaa 2438
      :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
QY 2113 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2143
      | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2439 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2469
      :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
```

RESULT 14

```
US-09-248-335-51
; Sequence 51, Application US/09248335
; Patent No. 6096504
; GENERAL INFORMATION:
; APPLICANT: MCGONIGLE, BRIAN
; APPLICANT: O'KEEFE, DANIEL
; TITLE OF INVENTION: PLANT GLUTATHIONE-S-TRANSFERASE ENZYMES
; FILE REFERENCE: CL-1128-A
; CURRENT APPLICATION NUMBER: US/09/248,335
; PRIOR FILING DATE: 1999-02-10
; EARLIER APPLICATION NUMBER: 08/924,759
; EARLIER FILING DATE: 1997-September-05
; NUMBER OF SEQ ID NOS: 74
; SOFTWARE: Microsoft Word Version 7.0A
; SEQ ID NO 51
; LENGTH: 967
; TYPE: DNA
; ORGANISM: maize
US-09-248-335-51
```

```
Query Match 2.6%; Score 56.4; DB 3; Length 967;
Best Local Similarity 68.4%; Pred. No. 3.3e-05;
Matches 78; Conservative 0; Mismatches 36; Indels 0; Gaps 0;
QY 2031 tacacattatttcttatttccctaatcattgaactcattcttaaaagaatc 2090
      | | | | | | | | | | | | | | | | | | | | | | | | |
```


100

GenCore version 4.5
Copyright (c) 1993 - 2000 Compgen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 25, 2002, 06:59:30 ; Search time 339.91 Seconds
(without alignments)
10829.527 Million cell updates/sec

Title: US-09-821-839-1
Perfect score: 2144
Sequence: 1 accgcacacagccacactctc.....aaaaaaaaaaaaaaaaa 2144

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SID55/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SID55/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SID55/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SID55/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2144	100.0	2144	22	AAD21310
2	1117.4	52.1	3870	22	AAD21312
3	123	5.7	3018	22	AAD21311
4	78.4	3.7	5204	24	ABL32901
5	63.4	3.0	545	22	AAH70126
6	63.4	3.0	1298	22	AAH72764
7	63.4	3.0	6564	24	ABL32123
8	63.2	2.9	1672	22	AAH92066
9	63.2	2.9	1701	21	AAZ64982

10	63.2	2.9	1701	22	AAH44128
11	62.8	2.9	268	22	AAH70080
12	62	2.9	216	22	AAH88314
13	62	2.9	216	22	AAH7565
14	62	2.9	255	22	AAH82171
15	62	2.9	705	22	AAH3177
16	61.8	2.9	246	22	AAH29131
17	61.2	2.9	215	22	AAH26432
18	61.2	2.9	6535	24	ABL32936
19	60.8	2.8	292	19	AAH71153
20	60.8	2.8	425	22	AAH60450
21	60.6	2.8	12177	24	ABL32651
22	60.4	2.8	337	22	AAH11578
23	60.4	2.8	1267	21	AAH77794
24	60.4	2.8	2497	21	AAH61197
25	60.4	2.8	5179	24	ABL33987
26	60.4	2.8	8946	24	ABL32911
27	60.4	2.8	15714	24	ABL33173
28	60.2	2.8	8920	15	AAH62934
29	60	2.8	1461	16	AAH01471
30	60	2.8	1461	22	AAH18736
31	59.8	2.8	7722	12	AAH12023
32	59.8	2.8	7753	15	AAH56826
33	59.6	2.8	443	8	AAH70773
34	59.6	2.8	3302	22	AAH34933
35	59.6	2.8	3380	22	AAH81743
36	59.4	2.8	768	22	AAH08093
37	59.4	2.8	5153	24	ABL32955
38	59.4	2.8	5464	24	ABL33916
39	59.4	2.8	11836	22	AAH45394
40	59.2	2.8	6070	24	ABL3679
41	59.2	2.8	6070	24	ABL34579
42	59	2.8	567	22	AAH89704
43	59	2.8	7346	22	ABL32345
44	58.8	2.7	612	22	AAH71471
45	58.8	2.7	245	22	AAH29057

ALIGNMENTS

RESULT	ID	Standard	CDNA	BP	Location/Qualifiers
1	AAD21310	standard	CDNA	2144	
AC	AAD21310				
DT	28-JAN-2002	(first entry)			
DE	Arabidopsis thaliana	SDS CDNA			
KW	Meiotic prophase I: chromosome 1: male sterile SDS mutant; apomixis; plant breeding; seed production; SDS protein; ss.				
OS	Arabidopsis thaliana				
FT	Key	Location/Qualifiers			
FT	CDS	98..1834			
FT	/*tag=	a			
FT	/product=	"SDS protein"			
PD	WO200174144-A1				
PD	11-OCT-2001				
PE	29-MAR-2001	2001WO-US09875			
PR	31-MAR-2000	2000US-193523P			
PA	(UNP-)	UNIV PENNSYLVANIA STATE			
FI	Ma H;				

Human PRO4063 (UNQ)
Human cervical can
Human digestive sy
Human colorectal c
Rat differential t
Human colon cancer
CDNA encoding for
Human breast cance
Human immune syste
3' nucleotide port
Human cancer agent
Human immune syste
Human breast cance
Human cancer assoc
Human cDNA encodin
Human immune syste
Human immune syste
Carbamoyl-phosphat
Mammalian mast cel
Rat mast cell func
Norwalk virus geno
Norwalk virus stra
Sequence encoding
Human colon cancer
Human differential
Human breast cance
Human immune syste
Chemically pretrea
Human immune syste
Human metastasis a
Human adult T-cell
Human immune syste
Human cervical can
CDNA encoding for

DR WPI: 2001-662939/76.
P-FSD8; AAE12998.

PT A new plant gene from Arabidopsis, designated SDS, mutations in which
PT are associated with inability to produce pollen, is useful for the
PT production of male sterile plants for plant breeding -

PS Claim 8; Page 39; 47pp; English.

XX The patent discloses novel plant gene from Arabidopsis thaliana,
CC designated SDS, which is associated with a failure to maintain
CC homologous attachment during meiotic prophase I. The SDS gene is
CC located on chromosome 1 and is useful in plant breeding to produce
CC male sterile SDS mutants and cloned progeny by apomixis where
CC meiosis is bypassed in seed production. The gene may also be
CC used as a probe to identify related genes in other plant species,
CC and to identify and isolate other genes of the meiosis regulatory
CC pathway. The present sequence is a cDNA encoding Arabidopsis thaliana
CC SDS protein.

XX Sequence 2144 BP; 649 A; 461 C; 467 G; 567 T; 0 other;

Query Match 100.0%; Score 2144; DB 22; Length 2144;

Best Local Similarity 100.0%; Pred. NO. 0; Mismatches 0; Indels 0; Gaps 0;

Matches 2144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 actgcatcagccactctctctagctctctgactaagaaacttccatttcaaaattcgaaatt 60
DB 1 actgcatcagccactctctctagctctctgactaagaaacttccatttcaaaattcgaaatt 60
QY 61 tctaatctctgattcgaagcttctgtaagagaaaatgaaagagatcgcatgagaa 120
DB 61 tctaatctctgattcgaagcttctgtaagagaaaatgaaagagatcgcatgagaa 120
QY 121 ttcaagagcaagccttgagcagagccgcttgccggaagaaagctcccgatcgagaa 180
DB 121 ttcaagagcaagccttgagcagagccgcttgccggaagaaagctcccgatcgagaa 180
QY 121 ttcaagagcaagccttgagcagagccgcttgccggaagaaagctcccgatcgagaa 180
DB 121 ttcaagagcaagccttgagcagagccgcttgccggaagaaagctcccgatcgagaa 180
QY 181 agcgcggaagagagacagatctctccgcttcttgaatacactctctgagcaaa 240
DB 181 agcgcggaagagagacagatctctccgcttcttgaatacactctctgagcaaa 240
QY 181 agcgcggaagagagacagatctctccgcttcttgaatacactctctgagcaaa 240
DB 181 agcgcggaagagagacagatctctccgcttcttgaatacactctctgagcaaa 240
QY 241 aatcggaagctctctgcttctgtaagatctctgctcgcgatttgcagtgatgacaagt 300
DB 241 aatcggaagctctctgcttctgtaagatctctgctcgcgatttgcagtgatgacaagt 300
QY 241 aatcggaagctctctgcttctgtaagatctctgctcgcgatttgcagtgatgacaagt 300
DB 241 aatcggaagctctctgcttctgtaagatctctgctcgcgatttgcagtgatgacaagt 300
QY 301 ttctctggtctcgagcagagtcgagagagctcgaaatccgaaagaaacttaattgaa 360
DB 301 ttctctggtctcgagcagagtcgagagagctcgaaatccgaaagaaacttaattgaa 360
QY 301 ttctctggtctcgagcagagtcgagagagctcgaaatccgaaagaaacttaattgaa 360
DB 301 ttctctggtctcgagcagagtcgagagagctcgaaatccgaaagaaacttaattgaa 360
QY 361 ggtagaagttcttaaacctggttaataatgtaagagagagatggttgaattcgaatttcg 420
DB 361 ggtagaagttcttaaacctggttaataatgtaagagagagatggttgaattcgaatttcg 420
QY 361 ggtagaagttcttaaacctggttaataatgtaagagagagatggttgaattcgaatttcg 420
DB 361 ggtagaagttcttaaacctggttaataatgtaagagagagatggttgaattcgaatttcg 420
QY 421 aagattacagaggtcttaacttaagcttacaagaagagagagagatgagatcgaaagt 480
DB 421 aagattacagaggtcttaacttaagcttacaagaagagagagagatgagatcgaaagt 480
QY 421 aagattacagaggtcttaacttaagcttacaagaagagagagagatgagatcgaaagt 480
DB 421 aagattacagaggtcttaacttaagcttacaagaagagagagagatgagatcgaaagt 480
QY 481 aaggaatcgcgtctggttgaattcgaattcgcgtgctggaatgaaggaatgaaagttaa 540
DB 481 aaggaatcgcgtctggttgaattcgaattcgcgtgctggaatgaaggaatgaaagttaa 540
QY 481 aaggaatcgcgtctggttgaattcgaattcgcgtgctggaatgaaggaatgaaagttaa 540
DB 481 aaggaatcgcgtctggttgaattcgaattcgcgtgctggaatgaaggaatgaaagttaa 540
QY 541 ggggaataaataaataaagcaagatgagatctcttccacagatccgattgtaacttcgc 600
DB 541 ggggaataaataaataaagcaagatgagatctcttccacagatccgattgtaacttcgc 600
QY 541 ggggaataaataaataaagcaagatgagatctcttccacagatccgattgtaacttcgc 600
DB 541 ggggaataaataaataaagcaagatgagatctcttccacagatccgattgtaacttcgc 600
QY 601 cggagatgctcacaagccggagattgaaattcgaatccgagaaataagagagcgagct 660
DB 601 cggagatgctcacaagccggagattgaaattcgaatccgagaaataagagagcgagct 660
QY 601 cggagatgctcacaagccggagattgaaattcgaatccgagaaataagagagcgagct 660
DB 601 cggagatgctcacaagccggagattgaaattcgaatccgagaaataagagagcgagct 660
QY 661 cgttctgcatatctcgtgagttgagttgtaagttcgaagttcgagagcgtttacccggagagc 720
DB 661 cgttctgcatatctcgtgagttgagttgtaagttcgaagttcgagagcgtttacccggagagc 720

DB 661 cgttctgcatatctcgtgagttgagttgtaagttcgaagttcgagagcgtttacccggagagc 720
QY 721 tgaataacgaagaattgaaatctcccaagccgagagcttctggaagctgattctctct 780
DB 721 tgaataacgaagaattgaaatctcccaagccgagagcttctggaagctgattctctctct 780
QY 781 tggatccgccaaggaattgaaagccgagcttgaaatagtcggaatgctctgactcgc 840
DB 781 tggatccgccaaggaattgaaagccgagcttgaaatagtcggaatgctctgactcgc 840
QY 841 ttgctcgaagaattctccgaaagaggttctgattctctcctcgaatgaaatgaaatgaa 900
DB 841 ttgctcgaagaattctccgaaagaggttctgattctctcctcgaatgaaatgaaatgaa 900
QY 901 accttcgaatatattcacaagatctccgacttcgaatctgattcgaatcgaatccgctcat 960
DB 901 accttcgaatatattcacaagatctccgacttcgaatctgattcgaatcgaatccgctcat 960
QY 901 accttcgaatatattcacaagatctccgacttcgaatctgattcgaatcgaatccgctcat 960
DB 901 accttcgaatatattcacaagatctccgacttcgaatctgattcgaatcgaatccgctcat 960
QY 961 ctcttcgactctgcaagcgaattctctgaaatctctcctcgaatctctcctattcaca 1020
DB 961 ctcttcgactctgcaagcgaattctctgaaatctctcctcgaatctctcctattcaca 1020
QY 1021 ttctcgtctctctgactctcgaatctcgaatctcgaatctcgaatctcgaatctcgaat 1080
DB 1021 ttctcgtctctctgactctcgaatctcgaatctcgaatctcgaatctcgaatctcgaat 1080
QY 1081 ttcttgatctctctgcaaggaagaattcgaatctcgaatctcgaatctcgaatctcgaat 1140
DB 1081 ttcttgatctctctgcaaggaagaattcgaatctcgaatctcgaatctcgaatctcgaat 1140
QY 1141 ggtggaagagagctatctcgaagctgaaaggaagaaagaaagaaagaaagaaagaaag 1200
DB 1141 ggtggaagagagctatctcgaagctgaaaggaagaaagaaagaaagaaagaaagaaag 1200
QY 1201 ctgttgtaagagcactgactgctccagagatggaacactgctcactcctcgtcgaagctc 1260
DB 1201 ctgttgtaagagcactgactgctccagagatggaacactgctcactcctcgtcgaagctc 1260
QY 1261 catcagatctcgaatctcgaatctcgaatctcgaatctcgaatctcgaatctcgaatctc 1320
DB 1261 catcagatctcgaatctcgaatctcgaatctcgaatctcgaatctcgaatctcgaatctc 1320
QY 1321 gttctcgaagatctgctcgttctggaatctcctcgaagcaagaaatcattcgaagcgaaag 1380
DB 1321 gttctcgaagatctgctcgttctggaatctcctcgaagcaagaaatcattcgaagcgaaag 1380
QY 1381 gactcgaatctcgaatctcgaatctcgaatctcgaatctcgaatctcgaatctcgaatctc 1440
DB 1381 gactcgaatctcgaatctcgaatctcgaatctcgaatctcgaatctcgaatctcgaatctc 1440
QY 1441 accttcaacagatctcgaatctcgaatctcgaatctcgaatctcgaatctcgaatctcgaat 1500
DB 1441 accttcaacagatctcgaatctcgaatctcgaatctcgaatctcgaatctcgaatctcgaat 1500
QY 1501 tgaagttggtgcaatgagagtgctggttcaagaagatcctcgaatctcgaatctcgaatctc 1560
DB 1501 tgaagttggtgcaatgagagtgctggttcaagaagatcctcgaatctcgaatctcgaatctc 1560
QY 1561 caaatctcgaatctcgttctggttcaatctcgaatctcgaatctcgaatctcgaatctcgaat 1620
DB 1561 caaatctcgaatctcgttctggttcaatctcgaatctcgaatctcgaatctcgaatctcgaat 1620
QY 1621 aaggaagcaaatctcgttctggttcaatctcgaatctcgaatctcgaatctcgaatctcgaat 1680
DB 1621 aaggaagcaaatctcgttctggttcaatctcgaatctcgaatctcgaatctcgaatctcgaat 1680
QY 1681 gccctcaactgtaagcagctgcaatctcgttctcgtcctgcatcgaacacacacacacacac 1740
DB 1681 gccctcaactgtaagcagctgcaatctcgttctcgtcctgcatcgaacacacacacacacac 1740
QY 1741 tgcataccaaagatgataaagttcgaatgaaagaaacagagagatgctcgaatg 1800
DB 1741 tgcataccaaagatgataaagttcgaatgaaagaaacagagagatgctcgaatg 1800

QY 1801 cgttaagagctcgcagctgttgcctggcagatgaagcaatcaaaaagaacaaaccctaa 1860
 |||
 Db 1801 cgttaagagctcgcagctgttgcctggcagatgaagcaatcaaaaagaacaaaccctaa 1860
 QY 1861 aaccagagacagctatctacgatacacaacacagctatcatcatctatctacacaaac 1920
 |||
 Db 1861 aaccagagacagctatctacgatacacaacacagctatcatcatctatctacacaaac 1920
 QY 1921 aaaaacagagtaagaataaagaacccctcctacagattatatacttaacgcagctgact 1980
 |||
 Db 1921 aaaaacagagtaagaataaagaacccctcctacagattatatacttaacgcagctgact 1980
 QY 1981 taattagagctcctagatatacacaattatagtcgacacattgtgtcgcctacacattta 2040
 |||
 Db 1981 taattagagctcctagatatacacaattatagtcgacacattgtgtcgcctacacattta 2040
 QY 2041 ttctctatttccctcaatcatctagactctcatattcttcttaaaagaatattctctgttt 2100
 |||
 Db 2041 ttctctatttccctcaatcatctagactctcatattcttcttaaaagaatattctctgttt 2100
 QY 2101 gaaaaaagaaaaaagaaaaaagaaaaaagaaaaaagaaaaa 2144
 |||
 Db 2101 gaaaaaagaaaaaagaaaaaagaaaaaagaaaaaagaaaaa 2144
 RESULT 2
 AAD21312 standard; DNA; 3970 BP.
 ID AAD21312 standard; DNA; 3970 BP.
 AC AAD21312;
 XX 28-JAN-2002 (first entry)
 DT 28-JAN-2002 (first entry)
 XX Arabidopsis thaliana SDS gene.
 DE Arabidopsis thaliana SDS gene.
 XX Melotic prophase I; chromosome 1; male sterile SDS mutant; apomixis;
 KM plant breeding; seed production; SDS gene; ds.
 XX Arabidopsis thaliana.
 OS Arabidopsis thaliana.
 XX WO200174144-A1.
 PN 11-OCT-2001.
 PD 29-MAR-2001; 2001WO-US09875.
 PE 31-MAR-2000; 2000US-193523P.
 PR (UNP-) UNIV PENNSYLVANIA STATE.
 PA Ma H;
 XX Ma H;
 PI WPI; 2001-662939/76.
 DR A new plant gene from Arabidopsis, designated SDS, mutations in which
 XX are associated with inability to produce pollen, is useful for the
 PT production of male sterile plants for plant breeding -
 PT Claim 23; Page 42-43; 47pp; English.
 PS The patent discloses novel plant gene from Arabidopsis thaliana,
 CC designated SDS, which is associated with a failure to maintain
 CC homologous attachment during meiotic prophase I. The SDS gene is
 CC located on chromosome 1 and is useful in plant breeding to produce
 CC male sterile SDS mutants and is useful in plant breeding to produce
 CC meiosis is bypassed in seed production. The gene may also be
 CC used as a probe to identify related genes in other plant species,
 CC and to identify and isolate other genes of the meiosis regulatory
 CC pathway. The present sequence is SDS gene from Arabidopsis thaliana.
 CC Sequence 3970 BP; 1228 A; 781 C; 738 G; 1223 T; 0 other;

Query Match 52.1%; Score 1117.4; DB 22; Length 3970;
 Best Local Similarity 99.9%; Pred. No. 4,1e-218;
 Matches 1118; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 actgcatcagccacccctcctcagctcctcgaactcaacattccatttcaaatcgaatt 60
 Db 510 actgcatcagccacccctcctcagctcctcgaactcaacattccatttcaaatcgaatt 569
 QY 61 tctaatctcagcttcaagcttctcgtacggaagaaatgaagagatctcgtatgagaa 120
 Db 570 tctaatctcagcttcaagcttctcgtacggaagaaatgaagagatctcgtatgagaa 629
 QY 121 ttcaagagcagagctcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 180
 Db 630 ttcaagagcagagctcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 689
 QY 181 acgacgagagagagacagacagacacacacacacacacacacacacacacacacacac 240
 Db 690 acgacgagagagagacagac 749
 QY 241 aatcgagctcctcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 300
 Db 750 aatcgagctcctcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 809
 QY 301 ttccctgtgtcgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 360
 Db 810 ttccctgtgtcgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 869
 QY 361 ggtagaagcttctcaaacctcgttctcaagtgagagagagagagagagagagagagag 420
 Db 870 ggtagaagcttctcaaacctcgttctcaagtgagagagagagagagagagagagagag 929
 QY 421 aagatctcagagctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 480
 Db 930 aagatctcagagctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 989
 QY 481 aagcaatcgcctcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 540
 Db 990 aagcaatcgcctcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 1049
 QY 541 gggaaataaataaataaataaataaataaataaataaataaataaataaataaataa 600
 Db 1050 gggaaataaataaataaataaataaataaataaataaataaataaataaataaataa 1109
 QY 601 cggacatgctcccaac 660
 Db 1110 cggacatgctcccaac 1169
 QY 661 cgttctcgtcatalctcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 720
 Db 1170 cgttctcgtcatalctcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 1229
 QY 721 tgatacgaagaagaatctgaatctcgaagcagcagcagcagcagcagcagcagcagc 780
 Db 1230 tgatacgaagaagaatctgaatctcgaagcagcagcagcagcagcagcagcagcagc 1289
 QY 781 tggatcggcgaagagatctgaagcagcagcagcagcagcagcagcagcagcagcagc 840
 Db 1290 tggatcggcgaagagatctgaagcagcagcagcagcagcagcagcagcagcagcagc 1349
 QY 841 ttgctctagaagaatctcgaagagctcgaatctcgaatctcgaatctcgaatctcgaat 900
 Db 1350 ttgctctagaagaatctcgaagagctcgaatctcgaatctcgaatctcgaatctcgaat 1409
 QY 901 acgttcagagatatactcaacagctatccgaactcgaatctcgaatctcgaatctcgaat 960
 Db 1410 acgttcagagatatactcaacagctatccgaactcgaatctcgaatctcgaatctcgaat 1469
 QY 961 cttctcgaacttgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1020
 Db 1470 cttctcgaacttgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1529

QY 1021 ttctgcctctctgtacctccagttcaaggacagttctgtagatccacagttcccaaga 1080
|||||
Db 1530 ttctgcctctctgtacctccagttcaaggacagttctgtagatccacagttcccaaga 1589
QY 1081 ttgtgacattcttgcgaggaagaattcactctgaatt 1119
|||||
Db 1590 ttgtgacattcttgcgaggaagaagaattcactctgaagt 1628

RESULT 3

AAD21311
ID AAD21311 standard; DNA: 3018 BP.

XX AAD21311;

XX 28-JAN-2002 (first entry)

XX Arabidopsis thaliana SDS promoter DNA.

KW Meiotic prophase I; chromosome 1; male sterile SDS mutant; apomixis;
XX plant breeding; seed production; SDS protein; ds.

XX Arabidopsis thaliana.

XX WO200174144-A1.

XX 11-OCT-2001.

XX 29-MAR-2001; 2001WO-US09875.

XX 31-MAR-2000; 2000US-193523P.

XX (UTPE-) UNIV PENNSYLVANIA STATE.

XX Ma H;

XX WPI; 2001-662939/76.

PT A new plant gene from Arabidopsis, designated SDS, mutations in which
PT are associated with inability to produce pollen, is useful for the
PT production of male sterile plants for plant breeding -

PS Claim 21; Page 41-42; 47pp; English.

CC The patent discloses novel plant gene from Arabidopsis thaliana,
CC designated SDS, which is associated with a failure to maintain
CC homologue attachment during meiotic prophase I. The SDS gene is
CC located on chromosome 1 and is useful in plant breeding to produce
CC male sterile SDS mutants and cloned progeny by apomixis where
CC meiosis is bypassed in seed production. The gene may also be
CC used as a probe to identify related genes in other plant species,
CC and to identify and isolate other genes of the meiosis regulatory
CC pathway. The present sequence is SDS promoter DNA from
CC Arabidopsis thaliana.

XX Sequence 3018 BP; 996 A; 529 C; 476 G; 1017 T; 0 other;

Query Match 5.7%; Score 123; DB 22; Length 3018;

Best Local Similarity 100.0%; Pred. No. 9.1e-16;

Matches 123; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 acgcatcagccacactctctagctctgactaagaacttcatttcaaaattcgaatt 60
|||||

Db 2896 acgcatcagccacactctctagctctgactaagaacttcatttcaaaattcgaatt 2955
|||||

QY 61 tctaatctcagttcagactctgtacgagagaaaaaagagagatcgagatagagaa 120
|||||

Db 2956 tctaatctcagttcagactctgtacgagagaaaaaagagagatcgagatagagaa 3015
|||||

QY 121 ttc 123
|||

Db 3016 ttc 3018
|||

RESULT 4
ID ABL32901/C
XX ABL32901 standard; DNA: 5204 BP.
XX ABL32901;

XX 26-MAR-2002 (first entry)

DE Human immune system associated gene SEQ ID NO: 874.

XX Human; immune system disease; cytosine methylation; antiaesthatic;
XX antiarteriosclerotic; antianemic; cytosatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmologic;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW gene; ds.

XX Homo sapiens.

XX WO200200928-A2.

XX 03-JAN-2002.

XX 02-JUL-2001; 2001WO-EP07537.

XX 30-JUN-2000; 2000DE-1032529.

XX 01-SEP-2000; 2000DE-1043826.

XX (EPIC-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;

XX WPI; 2002-130909/17.

PT Nucleic acid comprising fragment of chemically modified gene, useful
PT for diagnosis and treatment of diseases associated with abnormal
PT cytosine methylation -

PS Claim 1; SEQ ID NO 874; 32pp + Sequence Listing; German.

CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention.

XX Sequence 5204 BP; 1463 A; 72 C; 1014 G; 2655 T; 0 other;

Query Match 3.7%; Score 78.4; DB 24; Length 5204;

Best Local Similarity 53.2%; Pred. No. 1.2e-06;

Matches 166; Conservative 0; Mismatches 146; Indels 0; Gaps 0;

QY 1833 aagcaatcaaaaagaacaaaaccctaaaccagacacagatcactcgcataccaacac 1892
|||

Db 2507 AAAAAATTAATAAAAAAATCTTAATAATAATAAAAAAATTAATAAAAAAATCTTAATAATA 2448
|||

QY 1893 acaggttaccattaccattaccacaacaacaagaagtaagtaagaacacccctac 1952
|||

Db 2447 AAAAAATTAATAAAAAAATCTTAATAATAATAAAAAAATCTTAATAATAATAATAATA 2388
|||

QY 1953 agattatatacttaactgagctgactgaatgaatgactctagatataccaattagtcg 2012
|||

Db 2387 AATCTTTACATTTAATATTCTCTACTTAACAATATTAATTAATTAATTAATTAATTAATTA 2328
|||

QY 2013 caccattgtcgtcctacacacattattcttatttccctaatcattagactctca 2072
|||

Db	1543	ttctcttgctaatcttggaagatlaactcatctttaaataatattgctcaagatctaa	1602
Qy	2105	aaaaaaaaaaaaaaaaaaaaaaaaaaaaa	2144
Db	1603	aaaaaaaaaaaaaaaaaaaaaaaaaaaaa	1642
RESULT	9		
ID	AA264982		
XX	AA264982 standard; cDNA; 1701 BP.		
AC	AA264982;		
XX			
DT	05-APR-2000 (first entry)		
XX			
DE	Membrane-bound protein PRO1063 encoding CDNA.		
XX			
KW	Membrane-bound polypeptide; PRO polypeptide; LDL receptor; TIE ligand;		
RV	pharmaceutical; receptor immunoadhesin; gene mapping; ss.		
XX			
OS	Homo sapiens.		
XX			
PN	W09963088-A2.		
XX			
PD	09-DEC-1999.		
XX			
PF	02-JUN-1999;	99WO-US12252.	
XX			
PR	02-JUN-1998;	98US-0087607.	
PR	02-JUN-1998;	98US-0087609.	
PR	02-JUN-1998;	98US-0087759.	
PR	03-JUN-1998;	98US-0087827.	
PR	04-JUN-1998;	98US-0088021.	
PR	04-JUN-1998;	98US-0088025.	
PR	04-JUN-1998;	98US-0088028.	
PR	04-JUN-1998;	98US-0088029.	
PR	04-JUN-1998;	98US-0088030.	
PR	04-JUN-1998;	98US-0088033.	
PR	04-JUN-1998;	98US-0088167.	
PR	05-JUN-1998;	98US-0088202.	
PR	05-JUN-1998;	98US-0088212.	
PR	05-JUN-1998;	98US-0088217.	
PR	09-JUN-1998;	98US-0088655.	
PR	10-JUN-1998;	98US-0088722.	
PR	10-JUN-1998;	98US-0088730.	
PR	10-JUN-1998;	98US-0088734.	
PR	10-JUN-1998;	98US-0088738.	
PR	10-JUN-1998;	98US-0088740.	
PR	10-JUN-1998;	98US-0088741.	
PR	10-JUN-1998;	98US-0088742.	
PR	10-JUN-1998;	98US-0088810.	
PR	10-JUN-1998;	98US-0088811.	
PR	10-JUN-1998;	98US-0088824.	
PR	11-JUN-1998;	98US-0088863.	
PR	11-JUN-1998;	98US-0088865.	
PR	11-JUN-1998;	98US-0088858.	
PR	11-JUN-1998;	98US-0088861.	
PR	11-JUN-1998;	98US-0088876.	
PR	12-JUN-1998;	98US-0088990.	
PR	12-JUN-1998;	98US-0089105.	
PR	16-JUN-1998;	98US-0089440.	
PR	16-JUN-1998;	98US-0089512.	
PR	16-JUN-1998;	98US-0089514.	
PR	17-JUN-1998;	98US-0089532.	
PR	17-JUN-1998;	98US-0089538.	
PR	17-JUN-1998;	98US-0089598.	
PR	17-JUN-1998;	98US-0089599.	
PR	17-JUN-1998;	98US-0089600.	
PR	17-JUN-1998;	98US-0089653.	
PR	18-JUN-1998;	98US-0089601.	
PR	18-JUN-1998;	98US-0089907.	

[illegible]

PR 18-AUG-1998; 98US-0096960.
 PR 18-AUG-1998; 98US-0097022.
 PR 19-AUG-1998; 98US-0097141.
 PR 20-AUG-1998; 98US-0097218.
 PR 24-AUG-1998; 98US-0097661.
 PR 26-AUG-1998; 98US-0097951.
 PR 26-AUG-1998; 98US-0097952.
 PR 26-AUG-1998; 98US-0097954.
 PR 26-AUG-1998; 98US-0097955.
 PR 26-AUG-1998; 98US-0097971.
 PR 26-AUG-1998; 98US-0097974.
 PR 26-AUG-1998; 98US-0097978.
 PR 26-AUG-1998; 98US-0097979.
 PR 26-AUG-1998; 98US-0097986.
 PR 26-AUG-1998; 98US-0098014.
 PR 31-AUG-1998; 98US-0098525.
 PR 16-SEP-1998; 98US-0100634.
 PR 12-JAN-1999; 99US-0115565.

PA (GETH) GENENTECH INC.

PI Baker K, Chen J, Goddard A, Gurney AL, Smith V, Watanabe CK;
 PI Wood WI, Yuan J;

DR WPI: 2000-072883/06.
 DR P-PSDB: AAY66534.

XX Membrane-bound proteins and related nucleotide sequences -
 PS Claim 2; Fig 65; 822pp; English.

XX The invention provides membrane-bound PRO polypeptides and
 CC polynucleotides encoding them. The PRO sequences of the invention were
 CC identified based on extracellular domain homology screening. The PRO
 CC sequences have homology with proteins including LDL receptors, TIE
 CC ligands and various enzymes. The membrane-bound proteins and receptor
 CC molecules are useful as pharmaceutical and diagnostic agents. Receptor
 CC immunoadhesins, for instance, can be used as therapeutic agents to block
 CC receptor-ligand interactions. The membrane-bound proteins can also be
 CC employed for screening of potential peptide or small molecule inhibitors
 CC of the relevant receptor/ligand interaction. The PRO encoding sequences
 CC are useful as hybridization probes, in chromosome and gene mapping and in
 CC the generation of antisense RNA and DNA. PRO nucleic acid sequences
 CC will also be useful for the preparation of PRO polypeptides, especially
 CC by recombinant techniques.

XX Sequence 1701 BP; 557 A; 288 C; 381 G; 475 T; 0 other;

XX Query Match 2.9%; Score 63.2; DB 21; Length 1701;
 XX Best Local Similarity 77.0%; Pred. No. 0.0012;
 XX Matches 77; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

XX 2045 ttattcccatcattagactcattctatcttaaaagaatctctgttgtaa 2104
 XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 XX Db 1572 ttctcttgcttaattggaagattcaaccatttttaacaaattctgctaagattuaa 1631
 XX
 XX QY 2105 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2144
 XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 XX Db 1632 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1671

XX RESULT 10

XX AAF44128 standard; cDNA: 1701 BP.

XX AC AAF44128;

XX 02-APR-2001 (first entry)

XX Human PRO4063 (UNQ128) nucleotide sequence SEQ ID NO:114.

XX Human: secreted and transmembrane protein; PRO: cytosolic;

KW cell death; cancer; chromosomal mapping; gene mapping; tissue typing;
 KW diagnostic assay; ss.

XX Homo sapiens.

XX WO200073454-A1.

XX 07-DEC-2000.

XX 30-MAR-2000; 2000MO-US08439.

XX 02-JUN-1999; 99MO-US12252.

XX 23-JUN-1999; 99US-0141037.

XX 07-JUL-1999; 99US-0143048.

XX 20-JUL-1999; 99US-0144758.

XX 26-JUL-1999; 99US-0145698.

XX 28-JUL-1999; 99US-0146222.

XX 17-AUG-1999; 99US-0149396.

XX 15-SEP-1999; 99MO-US21090.

XX 08-OCT-1999; 99US-0158663.

XX 30-NOV-1999; 99MO-US28313.

XX 01-DEC-1999; 99MO-US28301.

XX 16-DEC-1999; 99MO-US30095.

XX 20-DEC-1999; 99MO-US30911.

XX 05-JAN-2000; 2000MO-US00219.

XX 06-JAN-2000; 2000MO-US00376.

XX 11-FEB-2000; 2000MO-US03565.

XX 18-FEB-2000; 2000MO-US04341.

XX 22-FEB-2000; 2000MO-US04414.

XX 24-FEB-2000; 2000MO-US04914.

XX 24-FEB-2000; 2000MO-US05004.

XX 02-MAR-2000; 2000MO-US05841.

XX 15-MAR-2000; 2000MO-US06884.

XX 20-MAR-2000; 2000MO-US07377.

XX (GETH) GENENTECH INC.

XX Ashkenazi AJ, Baker KP, Bolstein D, Desnoyers L, Eaton DJ;

XX Ferrara N, Fong S, Gerber H, Gerritsen ME, Goddard A, Godowski PJ;

XX Grimaldi CJ, Gurney AL, Kijavyn IJ, Napier MA, Pan J, Paoni NF;

XX Roy MA, Stewart VA, Tumas D, Watanabe CK, Williams PM, Wood WI;

XX Zhang Z;

XX WPI: 2001-032160/04.

XX P-PSDB: AAB65177.

XX PRO polynucleotides used to produce polypeptides used to target

XX bioactive molecules such as toxins, radiolabels or antibodies, to

XX specific cells, to cause targeted cell death -

XX Claim 2; Fig 65; 935pp; English.

XX The present invention describes human secreted and transmembrane PRO
 CC proteins. The PRO proteins have cytostatic activity. The PRO proteins
 CC can be used for targeted delivery of bioactive molecules, such as
 CC toxins, radiolabels or antibodies, that cause cell death. PRO nucleotide
 CC sequences, and their fragments, can be used as hybridization probes, in
 CC chromosomal and gene mapping, and in the generation of anti-sense RNA
 CC and DNA. They may also be used to produce transgenic animals which are
 CC used to develop and screen therapeutically useful reagents. The PRO
 CC nucleotide and protein sequence can be used for tissue typing and in
 CC treating cancer. Anti-PRO antibodies can be used in diagnostic assays.
 CC AAF44270 to AAF44470 represent PCR primers and hybridisation probes used
 CC in the isolation of human PRO sequences. AAF44087 to AAF44269 and
 CC AAB65154 to AAB65300 represent human PRO polynucleotide and protein
 CC sequences given in the exemplification of the present invention.

XX Sequence 1701 BP; 557 A; 288 C; 381 G; 475 T; 0 other;

XX Query Match 2.9%; Score 63.2; DB 22; Length 1701;
 XX Best Local Similarity 77.0%; Pred. No. 0.0012;

us-09-821-839-1.rng

[illegible]

ID	AAK88314	standard; cDNA; 216 bp
XX		
AC	AAK88314;	
XX		
DT	05-NOV-2001	(first entry)

DE	Human digestive system antigen coding sequence SEQ ID NO: 630.
33	

KW Human; digestive system antigen; gene therapy; cancer; appendicitis; ulcerative colitis; infection; tuberculosis; disease; chronic colitis;

KW digestive system disorder; Meckel's diverticulum; ss.

OS Homo sapiens.

PN W0200155314-A2

PD 02-AUG-2001.
yy

PF 17-JAN-2001; 2001WO-US01324.
XY

PR	31-JAN-2000; 2000US-0179065.
PR	04-FEB-2000; 2000US-0180628

PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186350

PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076

PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205575

PR 07-JUN-2000; 2000005-0209467
PR 28-JUN-2000; 2000015-0214885

PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0215647.

PR 07-JUL-2000; 20000US-02168880
PR 11-JUL-2000; 20000US-0217487

PR 11-JUL-2000; 2000US-0217496
PR 14-JUL-2000; 2000US-0218290

PR 26-JUL-2000; 2000US-0220963
PR 26-JUL-2000; 2000US-0220964

PR 14-AUG-2000; 2000US-0224518
PR 14-AUG-2000; 2000US-0224519

PR	14-AUG-2000; 2000US-0225213
PR	14-AUG-2000; 2000US-0225214

PR	14-AUG-2000; 2000US-0225266
PR	14-AUG-2000; 2000US-0225267

PR 14-AUG-2000; 2000US-0225268
PR 14-AUG-2000; 2000US-0225270

PR 14-AUG-2000; 2000US-0225757

PR 18-AUG-2000; 2000US-0226279
PR 22-AUG-2000; 2000US-0226581
PR 22-AUG-2000; 2000US-0226858
PR 22-AUG-2000; 2000US-0227183

PR	23-AUG-2000;	2000US-0227009
PR	30-AUG-2000;	2000US-0228924
PR	01-SEP-2000;	2000US-0229287

PR	01-SEP-2000;	2000US-0229344
PR	01-SEP-2000;	2000US-0229345
PR	05-SEP-2000;	2000US-0229509

```

PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 21-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234927.
PR 25-SEP-2000; 2000US-0234998.
PR 25-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241809.
PR 01-NOV-2000; 2000US-0241826.
PR 08-NOV-2000; 2000US-0246417.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249217.

PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

(PHUMAN) HUMAN GENOME SCI INC.
Rosen CA, Barash SC, Ruben SM.
WPI: 2001-502630/55.
P-PSDB: AAM92541.

PT Polynucleotides encoding digestive system antigens, useful for
PT diagnosing, treating, preventing and/or prognosing disorders of the
PT digestive system, particularly cancer and cancer metastases -
XX
XX
PS Claim 1; SEQ ID NO 630; 986bp; English.
XX
XX The present invention provides the protein and coding sequences of a
CC number of human digestive system antigens. These can be used in the
CC diagnosis, treatment and prevention of digestive system disorders,
CC including cancer, Meckel's diverticulum, bacterial or parasitic
CC infections, appendicitis, Hirschsprung's disease, chronic colitis or
CC ulcerative colitis. The present sequence is a cDNA encoding a digestive
CC system antigen of the invention.
XX
SQ Sequence 216 BP; 103 A; 24 C; 26 G; 60 T; 3 other;
XX

Query Match 2.9%; Score 62; DB 22; Length 216;
Best Local Similarity 62.5%; Pred. No. 0.0015;
Matches 95; Conservative 1; Mismatches 56; Indels 0; Gaps 0;

QY 1993 agtataccaattatagtcacacatttgctgcgtacacacatttctatttc 2052
Dy ||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 30 agtatagaactataaacaataatgatattttaaacattttacataagtaaa 89

QY 2053 cctaattcattagactcattcattcttaaaagaatatttccttcttgaaaaaaa 2112
Dy ||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 90 ctgcatctttagacataactacatttaaaataagcgcatacttttaaaaaaa 149

QY 2113 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2144
Dy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 150 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 101

RESULT 13
AA157565
ID AA157565 standard; cDNA: 216 BP.
XX
AC AA157565;
XX
DT 19-OCT-2001 (first entry)
XX
DE Human colorectal cancer antigen cDNA SEQ ID NO: 29.
XX

```

KM	Human; colorectal cancer; colorectal cancer antigen; gene therapy; ss.	
XX	Homo sapiens.	
OS		
XX		
PN	WO200155350-A1.	
XX		
PD	02-AUG-2001.	
XX		
PF	17-JAN-2001; 2001WO-US01350.	
XX		
PR	31-JAN-2000; 2000US-0179065.	PR 25-SEP-2000; 2000US-0234998.
PR	04-FEB-2000; 2000US-0180628.	PR 26-SEP-2000; 2000US-0235484.
PR	24-FEB-2000; 2000US-0184664.	PR 27-SEP-2000; 2000US-0235834.
PR	02-MAR-2000; 2000US-0186350.	PR 27-SEP-2000; 2000US-0235834.
PR	16-MAR-2000; 2000US-0189874.	PR 29-SEP-2000; 2000US-0236327.
PR	17-MAR-2000; 2000US-0190076.	PR 29-SEP-2000; 2000US-0236367.
PR	18-APR-2000; 2000US-0198123.	PR 29-SEP-2000; 2000US-0236368.
PR	19-MAY-2000; 2000US-0205515.	PR 29-SEP-2000; 2000US-0236369.
PR	07-JUN-2000; 2000US-0209467.	PR 29-SEP-2000; 2000US-0236370.
PR	28-JUN-2000; 2000US-0214886.	PR 02-OCT-2000; 2000US-0236802.
PR	30-JUN-2000; 2000US-0215135.	PR 02-OCT-2000; 2000US-0237037.
PR	07-JUL-2000; 2000US-0216647.	PR 02-OCT-2000; 2000US-0237038.
PR	07-JUL-2000; 2000US-0216880.	PR 02-OCT-2000; 2000US-0237039.
PR	11-JUL-2000; 2000US-0217487.	PR 02-OCT-2000; 2000US-0237040.
PR	11-JUL-2000; 2000US-0217496.	PR 13-OCT-2000; 2000US-0239935.
PR	14-JUL-2000; 2000US-0218290.	PR 13-OCT-2000; 2000US-0239937.
PR	26-JUL-2000; 2000US-0220963.	PR 20-OCT-2000; 2000US-0240960.
PR	26-JUL-2000; 2000US-0220964.	PR 20-OCT-2000; 2000US-0241221.
PR	14-AUG-2000; 2000US-0224518.	PR 20-OCT-2000; 2000US-0241785.
PR	14-AUG-2000; 2000US-0225213.	PR 20-OCT-2000; 2000US-0241786.
PR	14-AUG-2000; 2000US-0225214.	PR 20-OCT-2000; 2000US-0241787.
PR	14-AUG-2000; 2000US-0225266.	PR 20-OCT-2000; 2000US-0241808.
PR	14-AUG-2000; 2000US-0225267.	PR 20-OCT-2000; 2000US-0241809.
PR	14-AUG-2000; 2000US-0225270.	PR 20-OCT-2000; 2000US-0241826.
PR	14-AUG-2000; 2000US-0225447.	PR 01-NOV-2000; 2000US-0244617.
PR	14-AUG-2000; 2000US-0225757.	PR 08-NOV-2000; 2000US-0246474.
PR	14-AUG-2000; 2000US-0225758.	PR 08-NOV-2000; 2000US-0246475.
PR	14-AUG-2000; 2000US-0225759.	PR 08-NOV-2000; 2000US-0246476.
PR	18-AUG-2000; 2000US-0226279.	PR 08-NOV-2000; 2000US-0246477.
PR	22-AUG-2000; 2000US-0226681.	PR 08-NOV-2000; 2000US-0246478.
PR	22-AUG-2000; 2000US-0226686.	PR 08-NOV-2000; 2000US-0246523.
PR	22-AUG-2000; 2000US-0227188.	PR 08-NOV-2000; 2000US-0246524.
PR	23-AUG-2000; 2000US-0227009.	PR 08-NOV-2000; 2000US-0246525.
PR	30-AUG-2000; 2000US-0228924.	PR 08-NOV-2000; 2000US-0246526.
PR	01-SEP-2000; 2000US-0229287.	PR 08-NOV-2000; 2000US-0246528.
PR	01-SEP-2000; 2000US-0229343.	PR 08-NOV-2000; 2000US-0246532.
PR	01-SEP-2000; 2000US-0229344.	PR 08-NOV-2000; 2000US-0246609.
PR	01-SEP-2000; 2000US-0229345.	PR 08-NOV-2000; 2000US-0246610.
PR	05-SEP-2000; 2000US-0229509.	PR 08-NOV-2000; 2000US-0246611.
PR	05-SEP-2000; 2000US-0229511.	PR 08-NOV-2000; 2000US-0246613.
PR	06-SEP-2000; 2000US-0230437.	PR 17-NOV-2000; 2000US-0249207.
PR	06-SEP-2000; 2000US-0230438.	PR 17-NOV-2000; 2000US-0249208.
PR	08-SEP-2000; 2000US-0231242.	PR 17-NOV-2000; 2000US-0249209.
PR	08-SEP-2000; 2000US-0231243.	PR 17-NOV-2000; 2000US-0249210.
PR	08-SEP-2000; 2000US-0231244.	PR 17-NOV-2000; 2000US-0249211.
PR	08-SEP-2000; 2000US-0231413.	PR 17-NOV-2000; 2000US-0249212.
PR	08-SEP-2000; 2000US-0231414.	PR 17-NOV-2000; 2000US-0249213.
PR	08-SEP-2000; 2000US-0232080.	PR 17-NOV-2000; 2000US-0249214.
PR	12-SEP-2000; 2000US-0232081.	PR 17-NOV-2000; 2000US-0249215.
PR	12-SEP-2000; 2000US-0231968.	PR 17-NOV-2000; 2000US-0249216.
PR	14-SEP-2000; 2000US-0232397.	PR 17-NOV-2000; 2000US-0249217.
PR	14-SEP-2000; 2000US-0232398.	PR 17-NOV-2000; 2000US-0249218.
PR	14-SEP-2000; 2000US-0232399.	PR 17-NOV-2000; 2000US-0249244.
PR	14-SEP-2000; 2000US-0232400.	PR 17-NOV-2000; 2000US-0249245.
PR	14-SEP-2000; 2000US-0232401.	PR 17-NOV-2000; 2000US-0249264.
PR	14-SEP-2000; 2000US-0233064.	PR 17-NOV-2000; 2000US-0249265.
PR	14-SEP-2000; 2000US-0233065.	PR 17-NOV-2000; 2000US-0249297.
PR	21-SEP-2000; 2000US-0234223.	PR 17-NOV-2000; 2000US-0249300.
PR	21-SEP-2000; 2000US-0234274.	PR 01-DEC-2000; 2000US-0250160.
PR	25-SEP-2000; 2000US-0234997.	PR 01-DEC-2000; 2000US-0250391.
		PR 05-DEC-2000; 2000US-0251030.
		PR 05-DEC-2000; 2000US-0251988.
		PR 05-DEC-2000; 2000US-0256719.
		PR 06-DEC-2000; 2000US-0251719.
		PR 08-DEC-2000; 2000US-0251856.
		PR 08-DEC-2000; 2000US-0251858.
		PR 08-DEC-2000; 2000US-0251869.
		PR 08-DEC-2000; 2000US-0251989.
		PR 11-DEC-2000; 2000US-0254097.
		PR 05-JAN-2001; 2001US-0259678.

1
2
3
4
5